IDENTIFICATION OF MICRORNAS AND THIER TARGETS IN SWITCHGRASS, A MODEL CELLULOSIC BIOFUEL PLANT SPECIES

By

JESSICA ALLYN-BROOKER MATTS

Bachelor of Science in Biochemistry and Molecular Biology
Bachelor of Science in Biology

Oklahoma State University
Stillwater, OK
2005

Submitted to the Faculty of the
Graduate College of the
Oklahoma State University
in partial fulfillment of
the requirements for
the Degree of
DOCTOR OF PHILOSOPHY
December, 2010

IDENTIFICATION OF MICRORNAS AND THIER TARGETS IN SWITCHGRASS, A MODEL CELLULOSIC BIOFUEL PLANT SPECIES

Dissertation Approved:

Dr. Ramanjulu Sunkar

Dissertation Adviser

Dr. Ulrich Melcher

Dr. Ramamurthy Mahalingam

Dr. Yanqi Wu

Outside Committee Member

Dr. A. Gordon Emslie

Dean of the Graduate College

.

ACKNOWLEDGMENTS

First and foremost I would like to acknowledge my parents, Dr. Robert L. Matts and Janine Brooker for raising me to be inquisitive, and to aim for the stars. I also am grateful for the love and support from family, and the comic relief provided by my younger brother. I also am deeply grateful to the furry, four legged members of the family, especially Chewi and Pancakes for always being able to bring a smile to my face no matter how the day went.

I would also like to acknowledge my advisor Dr. Ramanjulu Sunkar, for his support and guidance over the years. I would also like to acknowledge the rest of my committee, Dr. Ulrich Melcher, Dr. Ramamurthy Mahalingram, and Dr. Yanqi Wu. I am especially grateful to Dr. Wu for providing not only the switchgrass seeds, but for allowing me to go to the greenhouses and collect tissue from mature plants. I would also like to acknowledge Dr. Wu's field technician Sharon, for caring for the plants, and the transportation to and from the greenhouses.

I would also like to acknowledge members of my lab, Dr. Guru Jagadeeswaran, for his help and guidance in building the small RNA libraries; Dr. Yongfong Li, Dr. Guosheng Li, and Li Zhang.

.

I would also like to acknowledge the core staff, Lisa Whitworth and Janet Rogers for the sequencing and help in troubleshooting when sequencing didn't work; Dr. Steve Hartson for the constant encouragement and brainstorming sessions when things weren't working. I would also like to acknowledge the office staff, for their help throughout the years.

I would finally like to acknowledge the friends who have lent an ear and a shoulder when needed, and those words of encouragement: Amanda Hale, Rachel Striver, Dr. Alisha Howard, Nicole Jones, Kristen Szabla, Lila Peal, Jacob Manjarrez, Liang Sun, Cherie Ognibene, and the rest of the graduate students in the department of Biochemistry and Molecular Biology.

TABLE OF CONTENTS

Chapter Page

LIST OF TABLES

Table Page

LIST OF FIGURES

Figure Page

LIST OF ABBRIVIATIONS

miRNA	microRNA
siRNA	small-interfering RNA
tasiRNA	trans-acting small-interfering RNA
MS media	Murashige & Skoog media
DEPC	Diethyl pyrocarbonate
PEG	Polyethylene glycol
RISC	RNA-induced silencing complex
dsRNA	double-stranded RNA
HD-ZIPIII	class III homeodomain-leucine zipper gene
ARF	auxin responsive factor
(SBP) transcription factors family	Squamosa Promoter Binding transcription factor
TCP factors transcription factor family	Teosinte branched1, Cycloidea, and PCF (TCP)
(AP2)-like transcription factor	Apetala-2 like transcription factor family

CHAPTER I

INTRODUCTION

In recent years, there has been a greater push for sustainable, alternative energy sources in regards to rising prices of fossil fuels and changing environmental conditions. Bio-fuels, solar, wind, and geothermal energies are some of the alternative forms of energies that are under consideration. Of these, biofuels is most promising form of alternative energy, with emphasis on production of ethanol from fermentable sugars. To date there are several plants being used for ethanol production, or are being evaluated for production of cellulosic ethanol (Table 1).

Table 1: Plant species and the type of biofuel produced from their sugars.

Plant Species	**In use or under evaluation**	**Annual or Perennial**	**Type of biofuel produced from sugars**
Sugarcane	In Use	Perennial	Ethanol
Corn	In Use	Annual	Ethanol
Switchgrass	Under evaluation	Perennial	Cellulosic Ethanol
Miscanthus	Under evaluation	Perennial	Cellulosic Ethanol
Sorghum	Under evaluation	Annual	Cellulosic Ethanol
Populus	Under evaluation	Perennial tree	Cellulosic Ethanol
Eucalyptus	Under evaluation	Perennial tree	Cellulosic Ethanol

In order to have a stable and continuous flow of biofuels, the future production of these fuels will require a constant supply of biomass, grown specifically for biofuel production. The production of biofuels from different biofuel plant species will require the ability to grow plants in regions that allow for the optimal survival and biomass production from the plant. Nevertheless, no single plant has both the optimal biomass production capabilities as well as optimal survival abilities everywhere, so more than one plant species will be needed. Therefore, increasing the understanding of the basic molecular biology of each plant species is necessary in order to optimize cellulosic biofuel production.

Both annual and perennial plant species are currently being evaluated for both cellulosic ethanol and ethanol production. The annual plants include wheat, alfalfa, sorghum, and waste products of corn and rice (stovers). The perennial plants include poplar and eucalyptus trees, and grasses such as switchgrass, *Miscanthus*, and *Brachypodium discantyon*. The major benefits for using perennial plants rather than annuals are: after the initial input of time and energy for the planting, perennials can re-grow after harvesting every year; within the US, switchgrass has been selected as dedicated perennial biomass producing plant species for biofuel production while the other plant species such *Miscanthus*, *Brachypodium*, *Populus*, etc. are under consideration.

Switchgrass

Switchgrass is a perennial, rhizomatous (with nodes) grass that is native to the prairies of North America (Figure 1). Over the past several years, switchgrass has emerged as one of the major biofuel plant sources in the United States, besides its utility as a forage crop (Schmer, *et al.* 2008; Keshwani and Chang, 2009). It is considered an ideal candidate for biofuel production for several reasons: It can (1) grow very tall (3-10 feet tall depending

on the ecotype/cultivar), (2) thrive well in marginal and waste lands, requiring very little fertilizer application, (3) also thrive in drought conditions, and (4) serve as a carbon sink, due to its large root system (Clark, 2002).

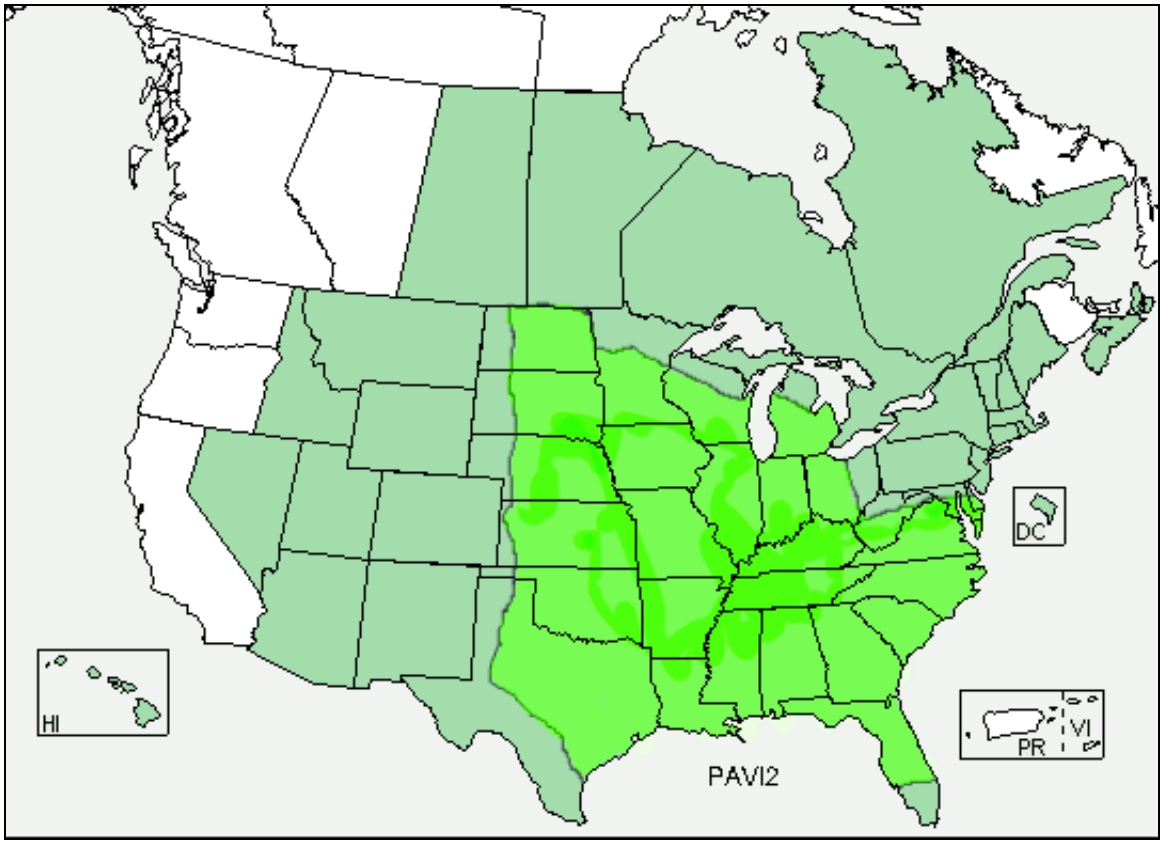


Figure 1. Distribution map of switchgrass in the United States and Canada. The lighter green region showcases the region of the United States that focuses on switchgrass for biofuel production. Modified from both http://plants.usda.gov/java/profile?symbol=PAVI2 and http://genomicscience.energy.gov/centers/map.jpg

Switchgrass is represented by two major ecotypes: upland (which are octoploid and found in the colder, northern climates), and lowland (which are tetraploid and found in the milder, dryer southern climates) (Bouton, 2007). With a wide geographic span, it is expected

that these different ecotypes have adapted to varied soil conditions found throughout the country. The lowland ecotypes seem to thrive more in soils that have a relatively high level of water moisture, whereas the upland ecotypes can thrive in soils that have water moisture levels that range from balanced to low (Figure 2) (Barney, *et al.* 2009).

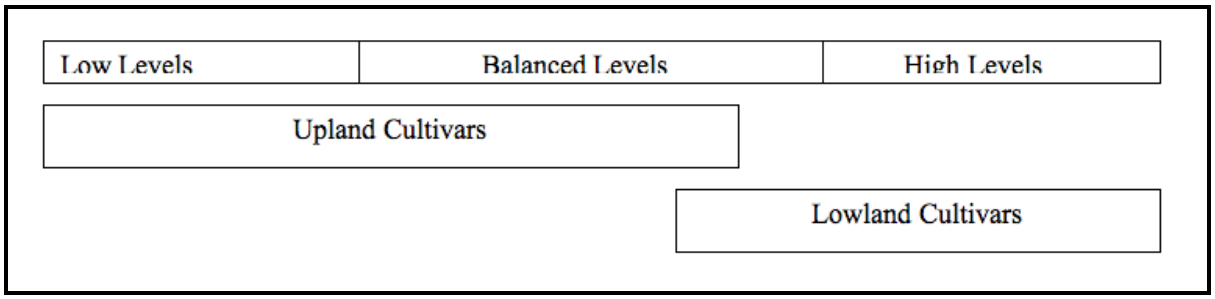


Figure 2. Basic classification of moisture levels found in soil (high, balanced, and low) and distribution of switchgrass cultivars relative to the soil moisture content.

Switchgrass is being grown currently as a forage material and for erosion control. Switchgrass plantations serve as shelter for numerous birds and small mammals, an ecological benefit of switchgrass under natural settings. Despite its increasing importance as an energy crop, little is known about the basic biology of the traits that control the utility of switchgrass as a biofuel crop. Identification of the gene regulatory processes and networks controlling plant biomass yield, nutrient uptake and assimilation and stress responses is an essential step forward to understand the gene regulatory processes in this important biofuel plant species. The findings could also lead to the rational design of strategies aimed at improving not only switchgrass biomass production but also other related biofuel plant species such as *Brachypodium*, *Miscanthus*, and others.

One of the primary objectives for scientists working on bioenergy production is to improve biomass accumulation of a biofuel plant species. Recent studies demonstrated that the transition from vegetative phase to reproductive phase can be blocked by manipulating the expression of microRNA156 in *Arabidopsis* which resulted in moderate delay in flowering, a severe decrease of apical dominance, and a prolonged vegetative phase (Schwab, *et al.* 2005). The combination of these traits led to a ten-fold increase in total leaf number in transgenic plants when compared to wild-type plants (Schwab, *et al.* 2005). Similar results were obtained in transgenic rice overexpressing miR156 (Liu et al., 2006). These results suggest that the identification of microRNAs (miRNAs) involved in the regulation of important plant characteristics such as phase change are attractive targets for improving biomass production.

Because miRNAs are involved in a myriad of biological functions, their identification in diverse plant species has been one of the most active research areas in recent years. To obtain better insight into the biological function of miRNAs in general, and individual miRNAs in particular, it is essential to identify all miRNAs that are expressed in a plant species. This endeavor is as important as mining genes that code for proteins. Such efforts have largely been focused on several model and crop plant species such as *Arabidopsis*, rice, *Populus, Medicago truncatula*, soybean, tomato, wheat and some other plant species. Thus far, miRNAs have not been identified in switchgrass. This thesis addresses the identification of conserved and novel miRNAs, determining their expression profile, and identification of their mRNA targets in switchgrass. The knowledge is likely to enhance the breadth of switchgrass molecular genetics and form a valuable resource for the large community of researchers working on switchgrass and other biofuel plant species.

CHAPTER II

LITERATURE REVIEW

The proper growth and development of a plant, its metabolism and stress responses, as well as numerous other functions, depends on the correct regulation of gene expression. Regulation of gene expression at the transcriptional level is dependent on the action of specific transcription factors that bind to conserved, *cis*-acting promoter elements, and is the most widely studied gene regulatory mechanism. Post-transcriptional gene regulation (mRNA decay and prevention of protein synthesis) was thought to be one of the critical mechanisms of gene regulation for normal growth and development and resistance to the adverse environmental conditions. However, the components that mediate this process are relatively unknown and only recently small RNAs (microRNAs and other endogenous siRNAs), which act as guide molecules in this process has been uncovered (Jones-Rhoades, *et al*. 2006; Mallory and Vaucheret, 2006; Sunkar, *et al*. 2007). This small RNA mediated regulation relies on specific RNA-RNA interactions that result in either target mRNA decay or suppression of the target mRNA protein production. In plants, these endogenous small RNAs can be divided into two main groups: microRNAs (miRNAs) and small-interfering RNAs (siRNAs), based on their biogenesis and function.

2.1: Small-interfering RNAs (siRNAs)

In recent years, large scale sequencing of small RNA libraries has revealed an unexpected diversity of endogenous siRNAs in plants (Llave *et al*., 2002; Sunkar and Zhu, 2004; Borsani *et al.*, 2005; Sunkar *et al*., 2005a; Rajagopalan *et al*., 2006; Fahlgren *et al*., 2007; Shukla *et al.*, 2008; Subramanian *et al.*, 2008). Interestingly, only a small fraction (1-2%) of the total small RNAs represented miRNAs, while the major fraction represented different classes of endogenous siRNAs such as tasiRNAs (trans-acting siRNAs), natsiRNAs (natural *cis*-acting siRNAs), heterochromatic siRNAs, and unidentified classes of small RNAs (Table 2).

Table 2. Classification of endogenous siRNAs in plants.

siRNA class	Location of origin	Biogenesis	Effector complex	Function
Trans-acting (TAS) siRNA	TAS loci	miR173 or miR390; DCL4	AGO1 or AGO7	mRNA cleavage
Natural cis-acting siRNA	Convergent gene pairs	DCL1	AGO??	mRNA cleavage
Heterochromatic associated siRNA	Heterochromatic loci	DCL3	AGO4/AGO6	DNA methylation & chromatin remodeling

siRNA biogenesis

Endogenous siRNAs are derived from the processing of long dsRNA into 21-24nt small RNAs (Waterhouse, et al. 2001; Plasterk, 2002; Axtell, ext al. 2006). dsRNAs are generated as a result of RNA-dependent RNA polymerase (RDR) activity on aberrant transcripts or read-troughs of the inverted repeats or antisense transcripts derived from the same loci. The resulting dsRNA is processed by the DCL (Dicer-like) family of

enzymes (DCL2, 3, and 4) to produce two classes of siRNAs: the 21-nt class and the 24-nt class.

siRNA function

The 21-nt siRNAs direct post-transcriptional silencing *via* mRNA degradation. This class can be referred to as either *cis*-acting siRNAs, if targeting the same loci in which they arose from, or as *trans-acting* siRNAs, because they regulate endogenous target mRNAs *in trans* (Peragine *et al.*, 2004; Vazquez *et al.*, 2004b; Allen *et al*., 2005). natsiRNAs are another sub-class of 21-nt endogenous siRNAs, which derive from a pair of sense and antisense transcripts in the cell (Borsani, *et al*. 2005). The second class, 24 nt siRNAs, is also referred to as heterochromatin-associated siRNAs (Hamilton and Baulcombe, 1999; Zilberman *et al.*, 2004). In *Arabidopsis*, these siRNAs are required for retroelement silencing through DNA and histone (H3K9) methylation (Chan *et al*., 2004; Zilberman *et al*., 2004; Tran *et al.*, 2005). While the functions of the repetitive element-associated 24 nt siRNA in heterochromatin formation (Chan *et al*., 2004; Zilberman *et al.*, 2004; Tran *et al*., 2005) and the 21 nt trans-acting siRNAs (Peragine *et al.*, 2004; Vazquez *et al.*, 2004b; Allen *et al.*, 2005; Williams *et al.*, 2005) in gene regulation involved in development (Lelandais-Briere, et al. 2010; Nag and Jack, 2010; Cho, et al. 2008; Boualem, et al. 2008; Liu, et al. 2007; Willmann and Pethig, 2007; Wu and Poethig, 2006) and the 21 nt natsiRNAs (Borsani et al., 2005; Jin et al., 2005; Katiyar-Agarwal et al., 2006) in stress responses are relatively well understood, the function of the natural *cis*-acting siRNAs are not well understood.

2.2: MicroRNAs

The term "microRNA" describes an abundant class of ~21-nt long non-coding RNA molecules, which can regulate the expression of protein coding genes at the post-transcriptional level in eukaryotes. Since the discovery of *lin-4 in C. elegns in 1993 (Lee et al., 1993),* the existence of these small non-coding RNA molecules have been discovered throughout the animal kingdom (Ambros, 2003; Rajewsky, 2006; Thatcher *et al*., 2008); fungi (Nakayashiki *et al*., 2006; Nakayashiki and Nguyen, 2008); the amoeba, *Dictyostelium discoideum* (Hinas *et al*., 2007); the single cell green alga, *Chlamydomonas reinhardtii,* (Zhao *et al*., 2007) and the plant kingdom (Barakat et al., 2007; Itaya et al., 2008; Subramanian et al., 2008; Carra et al., 2009; Jagadeeswaran et al., 2009b).

MicroRNA biogenesis

MicroRNAs are transcribed by RNA polymerase II, and the primary miRNA (pri-miRNA) transcript is subsequently capped, spliced, and poly-adenylated (Kurihara and Watanabe, 2004; Lee *et al*., 2004; Borsani *et al*., 2005; Xie *et al*., 2005; Kurihara *et al*., 2006; Lu *et al*., 2006). The primary miRNA transcript has the ability to adopt a hairpin like secondary structure, which can be processed into miRNA:miRNA* (miRNA and miRNA-star) duplex by the action of Dicer-like 1 (DCL1), a double-stranded RNA-binding endonuclease. Two other nuclear-localized proteins hyponastic leaf 1 (HYL1, another dsRNA binding protein), and serrate (SE, a zinc finger protein), assist the DCL-1 in processing the primary miRNA into the miRNA:miRNA* duplex (Liu *et al*., 2005; Yang *et al*., 2006; Dong *et al*., 2008; Laubinger *et al*., 2008). Additionally, HYL1 and

RNA cap-binding protein also implicated in releasing the miRNA:miRNA* duplex from the hairpin-like structure (Yu et al., 2008). Another nuclear-localized protein that is also essential for miRNA accumulation is the *Arabidopsis* Hua Enhancer 1 (HEN1) protein, a methyltransferase. HEN1 methylates the 3' end of each strand of the duplex (Li *et al.*, 2005; Yang *et al.*, 2006). The methylation of the miRNA:miRNA* duplex prevents the 3' end uridylation (addition of an oligoU tail to the 3' ends of both the strands of the duplex). Uridylation of miRNAs might interfere with the function of the miRNA (ability to enter the RISC complex) and most importantly uridylation can serve as a signal for rapid degradation of the miRNA duplex (Li *et al.*, 2005; Ibrahim *et al*., 2010). The nuclear localization of DCL1, HYL1, SE, and HEN1 suggests that miRNA biogenesis in plants takes place within the nucleus (Figure 3). The processed mature miRNA:miRNA* duplex will be subsequently exported to the cytosol via HASTY5, a plant ortholog of exportin5 (Bollman *et al*., 2003; Park *et al*., 2005).

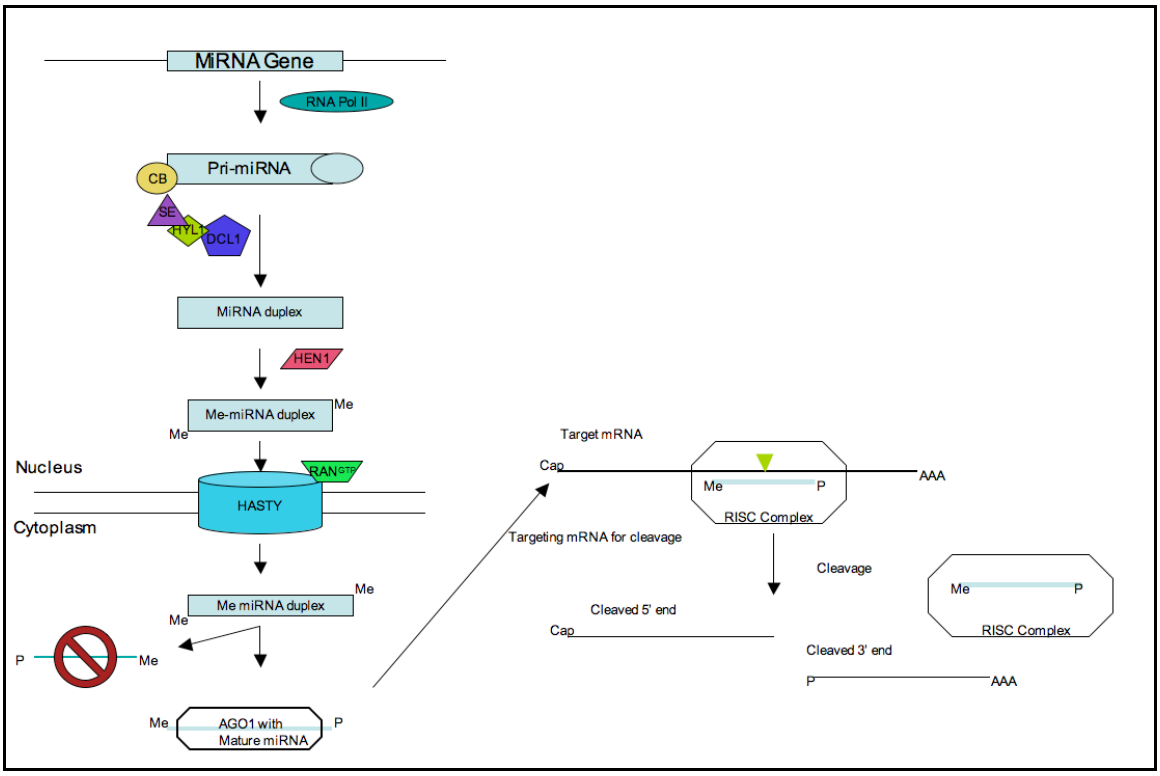


Figure 3. The biogenesis and function of miRNAs in plants. (Modified from Jones-Rhoades, et al. *Annual Review of Plant Biology*. 57:19-53, 2006).

Conserved and non-conserved miRNAs in plants

On the basis of miRNA conservation, plant miRNAs are classified into conserved and non-conserved miRNAs. To date, twenty-one miRNA families (miR156, miR159, miR160, miR162, miR164, miR165/166, miR167, miR168, miR169, miR170/171, miR172, miR319, miR390, miR393, miR394, miR395, miR396, miR397, miR398, miR399 and miR408) are conserved between monocot and dicot plants (Jones-Rhoades *et al*., 2006). Eight of these miRNA (miR156, miR160, miR166, miR168, miR169, miR170, miR172, and miR390) families are conserved even in primitive land plants such as *Physcomitrella* and *Selaginella*, (Axtell and Bartel, 2005; Axtell *et al.*, 2007). Recently developed high-throughput next generation sequencing technologies such as

MPSS (massively parallel signature sequencing), 454 (pyrosequencing), and SBS (sequencing-by-synthesis), have enabled the identification of near-completed sets of miRNAs in several plant species (Sunkar and Zhu, 2004; Arazi *et al*., 2005; Sunkar *et al*., 2005b; Sunkar *et al*., 2005a; Talmor-Neiman *et al*., 2006a; Talmor-Neiman *et al*., 2006b; Fahlgren *et al.*, 2007; Addo-Quaye *et al*., 2008; Lu *et al.*, 2008; Sunkar *et al*., 2008; Zhu *et al.*, 2008). These deep sequencing studies have revealed that the miRNA component of plants is comprised of lineage-specific and species-specific miRNA families, in addition to the ~21 well conserved families. Within the lineage-specific miRNAs some are broadly conserved whereas some others are conserved in closely related plant species. For instance, most examined dicotyledonous plants to date are found to encode for miR403, while its' counterpart could not be identified in monocots (Sunkar and Zhu, 2004; Sunkar and Jagadeeswaran, 2008). Similarly, miR396d, miR437, and mIR444, appear to exist in all most all monocots examined to date (Sunkar *et al*., 2005b; Lu *et al*., 2008; Sunkar and Jagadeeswaran, 2008), but not in dicots. A few other well-established examples are miR2119, miR2119 and miR2199 that were found to be conserved in closely related legumes but not in Arabidopsis, rice, *Populus*, *Sorghum*, and other plant species whose genomes are known (Jagadeeswaran et al., 2009). Deep sequencing efforts also revealed the existence of several species-specific miRNAs in Arabidopsis, rice, *Populus*, *Physcomitrella*, *Medicago truncatula* (Sunkar et al., 2005b; Pilcher et al., 2007; Sunkar et al., 2008; Carra et al., 2009; Jagadeeswaran et al., 2009b). The existence of both lineage-specific and species-specific miRNAs implies a complex post-transcriptional regulatory network operating in plants, and that these species-specific miRNAs could have specific roles in unique pathways not shared among all plants.

microRNA function

Once the miRNA:miRNA* complex has entered the cytoplasm, the duplex disassociates and the guide strand is incorporated into the RISC (RNA interference silencing complex), containing an Argonaute (AGO) protein, while the miRNA* is either degraded or accumulates at very low levels. This preferred RISC assembly with one strand of the miRNA duplex is referred to as the asymmetric assembly of RISCs (Schwarz et al., 2003; Khvorova et al., 2003). Interestingly, both in animals and plants some miRNA* species accumulate to detectable levels (Jagadeeswaran et al., 2009). A few such miRNA* have been reported to regulate the gene expression of their target mRNAs in animals (Okamura *et al*., 2008). However, whether the miRNA* is functional or not remains unknown in plants. In plants, guided by the miRNA, AGO1 catalyzes the cleavage between the 10^{th} and 11^{th} nt in the complementary region of the target mRNA. However, the mode of miRNA-guided target regulation differs in animals, in which the seed region (the seven nucleotides from 2 to 8 from the 5' end), determines the target specificity, and has multiple target sites in its 3'-UTRs. This combination (a small seed region and multiple target sites in the 3'-UTR) leads more to translational repression rather than cleavage of animal mRNA targets (Ambros, 2004; Millar and Waterhouse, 2005; Brodersen and Voinnet, 2009; Chekulaeva and Filipowicz, 2009).

Our current knowledge about the regulatory roles of miRNAs and their targets points to fundamental functions in various aspects of plant development including auxin signaling, meristem boundary formation and organ separation, leaf development and polarity, lateral root formation, transitions from both juvenile-to-adult vegetative phase

and vegetative-to-flowering phase, floral organ identity, and reproduction (reviewed in Jones-Rhoades *et al*., 2006; Mallory and Vaucheret, 2006).

2.3 trans-acting siRNAs (tasiRNA) biogenesis and function

In Arabidopsis, the action of certain miRNAs such as miR173, miR390, and miR828 is essential for the generation of trans-acting siRNAs (tasiRNAs). Initially, the miRNA173/390/828 containing AGO complex cleaves the non-coding transcript, which then serves as the template for the dsRNA biogenesis through the action of RdRP (Vaucheret, 2005).

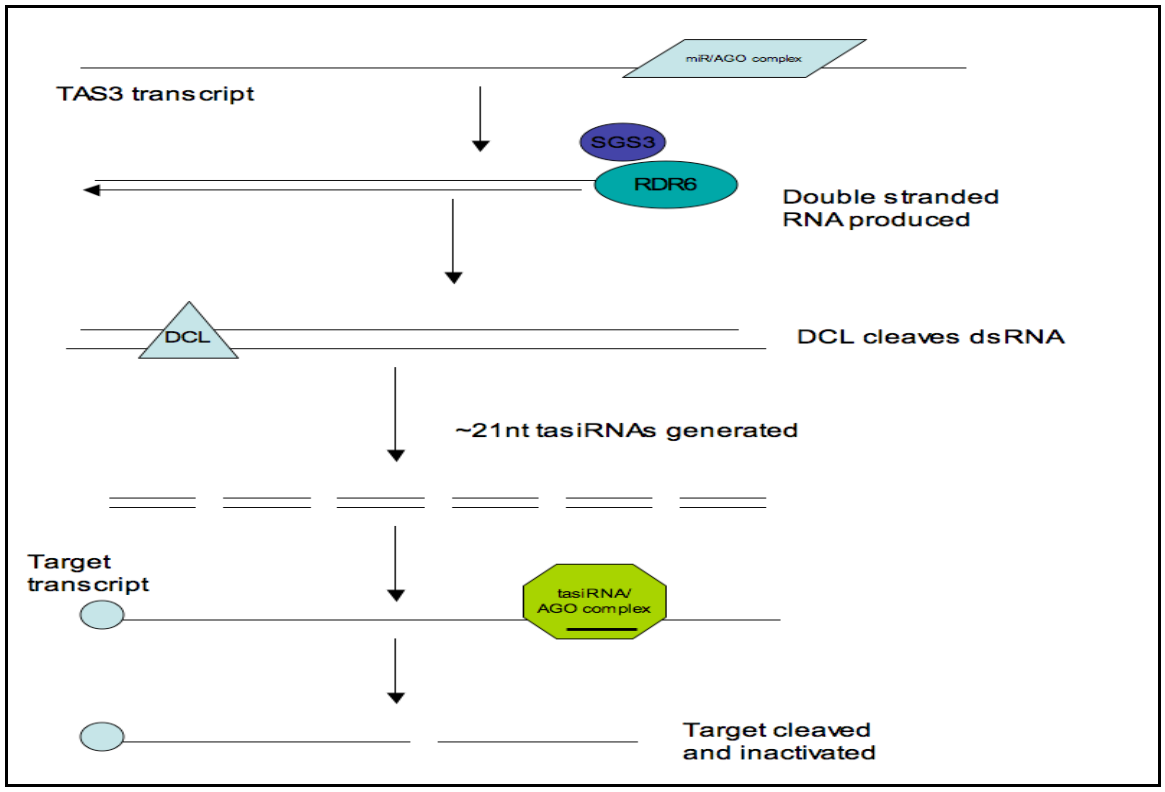


Figure 4. Biogenesis and function of trans-acting siRNAs. Modified from Xie, *et al. Proc. Natl. Acad. Sciences USA*, 102(36): 12984-12989, 2005

miR390 has been shown to interact with Argonaute7 to produce the TAS3 siRNAs from the TAS3 locus in *Arabidopsis* (Montgomery e*t al*., 2008a). These tasiRNAs in turn target and regulate the expression of auxin response factors ARF3 and ARF4 genes in plants. ARF3 and ARF4 regulation by Tas3siRNAs is important for lateral organ development in Arabidopsis (Fahlgren *et al*., 2006). There are three other TAS loci in *Arabidopsis*. TAS1 and TAS2 are targeted by miR173 acting with AGO1 (Montgomery *et al*., 2008b), while TAS4 is targeted by miR828 (Allen *et al*., 2005). The targets of TAS1 and TAS2 include several pentatricopeptide repeat proteins (PPRs) (Montgomery *et al*., 2008b), while TAS4 targets MYB75 (Hsieh *et al*., 2009).

2.4 Role of miRNAs in plant stress responses

miRNAs have been shown to play important roles in response to both nutrient deprivation and other environmental stresses in addition to their roles in plant growth and development (Jones-Rhoades and Bartel, 2004; Sunkar and Zhu, 2004; Fujii *et al*., 2005; Sunkar *et al*., 2006; Sunkar, 2010). For example, miR395 could not be detected in plants grown on medium containing optimal levels of sulfate, but is induced when sulfate levels are depleted in the medium. miR395 is targeting three ATP sulfurylases (APS1, APS3 and APS4), enzymes that catalyze the first step of inorganic sulfate assimilation, and *APS1* encoding a sulfate transporter. *APS1* mRNA is negatively correlated with the miR395 levels under low-sulfate stress (Jones-Rhoades and Bartel, 2004). Similarly, miR399 could not be detected in plants grown on medium containing optimal levels of phosphate, but is induced when sulfate levels are depleted in the medium (Fujii, *et al.* 2005). By contrast to the induction of specific miRNAs under stress conditions, the

opposite effect (down-regulation of miRNA) has been also reported. For instance, miR398, which targets two superoxide dismutase (*CSD1* and *CSD2*) transcripts, has its own transcription down-regulated to facilitate the increased expression of superoxide dismutase transcripts (Sunkar, *et al*. 2006). Besides, the observation that several miRNAs critical for growth and development are altered during stress implies that cessation of plant development during stress is attained, at least in part, by up-regulating miRNAs that suppress growth and development by down-regulating the target genes, which are positive regulators of growth and development or down-regulating miRNAs to up-regulate target genes that act as negative regulators of plant growth and development or a combination of both (Sunkar, 2010). Overall, it has been well established that miRNAs are an integral part of stress response regulatory networks in plants.

CHAPTER III

MATERIALS AND METHODS

In general, miRNAs can be identified using two different approaches, either using computational approach or direct cloning of small RNAs. Application of computational approach can only predict conserved miRNAs but not species-specific miRNAs. Here, we used both approaches for identification of miRNAs in switchgrass.

3.1: Material

Switchgrass, cultivar Alamo was used in this study and the seeds were a kind gift from Dr. Yanqi Wu, Department of Plant and Soil Science, Oklahoma State University.

3.2: Growth Conditions

Seedlings were grown in growth chambers with a 16-/8-h day/night cycle (with a light intensity of 1,050-1,250 lux) at 21°C for 8-12 weeks, and adult plants were grown in a greenhouse, maintained by Dr. Yanqi Wu's lab.

3.3: Tissue Selection

Eight different tissue samples were selected for analyzing the temporal expression of miRNAs in switchgrass. These include 'lower leaves' from seedlings and adult

adult plants (three to four lowest leaves), 'upper leaves' from seedlings and adult plants (three to four uppermost leaves), 'stems' from the seedlings and adult plants. Additionally, 'inflorescence' from the adult plants and 'roots' from the seedlings were utilized. After harvesting, the tissues were snap-frozen in liquid nitrogen and stored at -80°C until RNA extractions were performed.

3.4: Nutrient Stress treatments

Seeds were germinated on wet vermiculite, and were grown in growth chambers with a 16-/8-h day/night cycle (with light intensities of 1,050-1,250 lux) at 21°C for 3 weeks and then the seedlings were transferred to 96-well PCR plates with holes (bottoms were cut) in their wells. The seedlings were grown in the presence of control media (MS with all supplements) for 4 weeks, and then transferred to modified MS media containing different levels of sulfate or phosphate (0.02, 0.2, or 2.0mM) for 5 days. The media was replaced daily. After treatment, seedlings were collected and frozen in liquid nitrogen, and stored at -80°C until RNA was extracted.

3.5: Total RNA isolation

Total RNA was isolated from selected tissues using the Trizol method (Invitrogen, USA). In short, the tissue was ground in liquid nitrogen to a fine powder and transferred to a 50ml conical tube containing Trizol. The tube was weighed and additional Trizol was added to ensure that every 100 mg of ground tissue was resusupended in 1 ml of Trizol. The sample was incubated at room temperature for 5 minutes, and then centrifuged at 13,0000g for 5 minutes. The soluble fraction was then

transferred to another 50ml conical tube, and 200µl chloroform was added for every one ml Trizol. The sample was vigorously mixed for 2 minutes at and then centrifuged at 13,000g for 10 minutes. The top aqueous layer was removed to a fresh tube; 500µl isopropanol/ml Trizol was added. The sample was mixed and placed at -20°C overnight to allow for RNA precipitation. The sample was then centrifuged at 13,000g for 20 minutes to pellet the RNA. Once the liquid was carefully removed without disturbing the pellet, the pellet was washed with 500µl of 80% ethanol by centrifuging at 13,000g for 5 min and then the ethanol was removed and the pellet was air-dried briefly. The pellet was then resusupended in 500µl of DEPC treated water. Total RNA concentration was determined by using a nanodrop ND-1000 spectrophotometer, and the RNA integrity was determined by resolving on a 2% agarose gel.

3.6: Low-Molecular Weight (LMW) RNA isolation

A low-molecular weight RNA fraction was isolated from the total RNA by precipitating high molecular weight RNAs (mRNAs and rRNAs) with 10% polyethylene glycol (mol. wt 8000) and 0.5 M NaCl (4°C for 30 min) and centrifugation (13,000g for 30 min) (Hamilton et al., 2002). The supernatant was collected and the low molecular weight RNAs including miRNAs were precipitated using 3 volumes of cold ethanol and incubating at –20°C for overnight. (Hamilton et al. 2002). The next day the samples were centrifuged again at 13,000g for 30 minutes at 4°C. The supernatant was carefully removed and the pellet was washed with 0.5ml of 80% ethanol by centrifuging at 13,000g for 5 minutes at 4°C. Then the ethanol was carefully removed and the pellet was briefly

air-dried and resusupended in 100µl of DEPC-treated water. Aliquots of 2µl were used to determine the RNA concentration using a nanodrop ND-1000 spectrophotometer.

3.7: Small RNA Library Construction

Three-month old seedlings, emerging tillers and inflorescence from the adult plants were collected, snap frozen in liquid nitrogen and stored at -80°C and used for total RNA isolation. Total RNA was extracted using the Trizol reagent (Invitrogen, USA) following the manufacturer's instructions. Total RNA was then size-fractionated on a denaturing 15% polyacrylamide/8M urea gel at 220V/cm for 1-2 hours, along with ^{32}P labeled 21 and 24-nt RNA oligos. These labeled RNA oligos served as markers for isolating the small RNAs of 21 to 24-nt in size. The gel piece around the 21-24nt was excised and the small RNAs were eluted by gentle shaking of the small gel pieces in 3 volumes of 0.3M NaCl at 4°C overnight. The small RNAs were then precipitated with the addition of 100% ethanol and incubated at -80°C overnight. The samples were then spun at 13,000g for 30 minutes at 4°C, and then the pellet was resusupended in 6µl of DEPC-treated water. An RNA adaptor was then ligated to the 5' ends of the isolated small RNAs using T4 RNA ligase. The ligation reaction was carried out at 37°C for 1 to 2 hrs. The 5'-adaptor ligated small RNAs were again size fractionated using a denaturing 15% polyacrylamide gel, and the samples were again isolated, precipitated, and resusupended in 6µl DEPC-treated water. A 3'RNA adaptor was then ligated to the 3' ends of the small RNAs using T4 RNA ligase at 37°C for 1 to 2 hrs. A reverse transcription reaction was performed using the RT primer (454 library, AAGGATGCGGTTAAA, CAAGCAGAAGACGGCATACGA, Solexa), A subsequent

PCR reaction was then performed using a forward primer (454 library, TACTAATACGACTCACTAAA; AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA, Solexa library) and a reverse primer (454 library, AAGGATGCGGTTAAA; CAAGCAGAAGACGGCATACG, Solexa library). The final PCR product was then run on a 3% low-melting agarose gel along with a 25bp DNA ladder, which served as a marker. The PCR products corresponding to the expected final size (93-96bp) were isolated and purified. A small aliquot (1-2µl) of the PCR products was used for cloning into pGEM-easy T vector (Promega, WI USA) in order to check the quality of the library. After satisfactory determination of the quality of the library, the PCR products were sequenced using pyrosequencing or sequencing-by-synthesis technologies. Figure 5 shows a schematic diagram for the construction of the libraries.

Adapters and primers used for 454 small RNA library:

5': 5'-tactaatacgactcactAAA-3'; uppercase, RNA; lowercase, DNA
3': 5'-pUUUaaccgcatccttctcx-3'; uppercase, RNA; lowercase, DNA; p, phosphate; x, inverted deoxythymidine. RNA/DNA chimeric oligonucleotide adapters
Forward PCR primer: TACTAATACGACTCACTAAA
Reverse PCR primer: AAGGATGCGGTTAAA

Adapters and primers used for Solexa small RNA library

5’ RNA adaptor: 5’-GUUCAGAGUUCUACAGUCCGACGAUC
3’ RNA adaptor: 5’-P-UCGUAUGCCGUCUUCUGCUUGUidT
PCR primer 1:
AATGATACGGCGACCACCGACAGGTTCAGAGTTCTACAGTCCGA
PCR primer 2: CAAGCAGAAGACGGCATACG

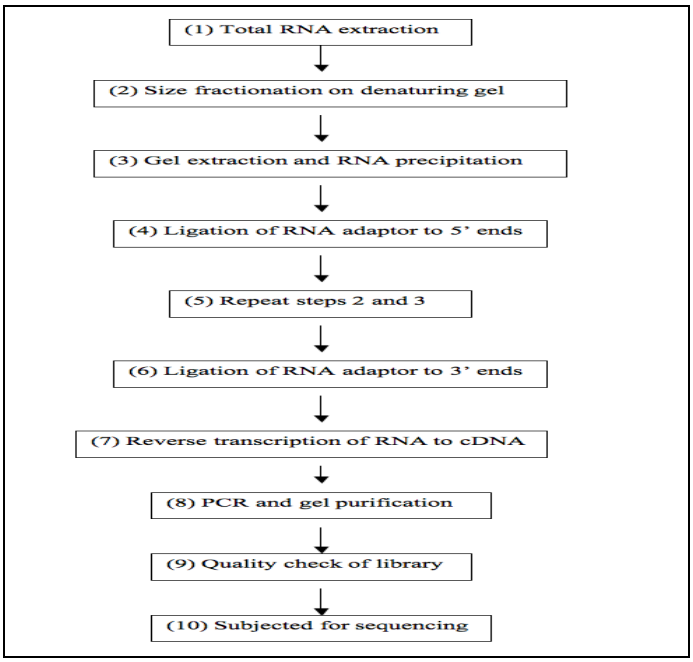


Figure 5. Schematic presentation of construction of a small RNA library.

3.8: Computational approach for predicting miRNA in switchgrass

A computational approach has been successful in identification of conserved miRNAs in diverse plants and animals (Sunkar and Jagadeeswaran 2008; Zhang, *et al*. 2006; Wang and El Naqa, 2008; Rajewsky, 2006). The basis for computational identification of miRNAs, is the conservation of mature miRNA sequence coupled with the predictable secondary structure for miRNA precursors (Ambros, *et al*. 2003). For identification of a "complete set" of conserved miRNAs, the availability of a complete genome sequence is a prerequisite. If the complete genome is unavailable, the available large genomic fragmented data in the form of GSS (genomic survey sequences), WGS (whole-genome shotgun reads), HTGS (high throughput genomic sequences) and NR (non-redundant nucleotide sequences) have been used for identification of conserved miRNAs in several plant species. Analysis of ESTs also aided in identification of several conserved miRNAs from diverse plant species (Sunkar and Jagadeeswaran, 2008; Zhang, *et al*. 2006). Currently no GSS, WGS, HTGS sequences are available for switchgrass, thus, we have to rely on ~436,535 switchgrass ESTs that were deposited at NCBI's database for the identification of conserved miRNAs. Thus, the number of ESTs available in the database

is a major limitation for identification of complete set of conserved miRNAs in switchgrass.

Selection of mature miRNA

The mature miRNA sequences from Arabidopsis and rice were downloaded from the miRNA database (http://www.mirbase.org) for the identification of conserved miRNAs. For identification of monocot specific miRNAs, miRNA sequences from rice were used.

Parameters for BLASTn searches and secondary structure predictions

Conserved miRNAs are highly identical in sequence. Of the 21 nucleotides, either all 21 nucleotides are conserved or the sequence will differ by only one or two nucleotides between diverse plant species to identify miRNA homologs of conserved miRNAs in switchgrass. The mature miRNA sequences were used as a query in blast searches against NCBI's EST database. We used NCBI BLASTN to find orthologous/paralogous miRNA sequences matching at least 18 nucleotides and leaving 3 nucleotides for possible sequence variations in different plant species. Hits among the ESTs with sense orientation (plus/plus orientation) with 0-3 mismatches were candidates for conserved miRNAs. If multiple hits were found, such ESTs were aligned and unique sequence was extracted and used for predicting secondary structure to the miRNA precursor. The flanking region of 300 bp upstream and downstream to the mature miRNA sequence was used for fold-back structure predictions using mFOLD (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) program. The obtained secondary structures were analyzed and compared to secondary structures deposited in the miRNA database (http://www.mirbase.org) in order to verify that the miRNA was located on the same arm

as it's counterpart in different plant species (Figure 6).

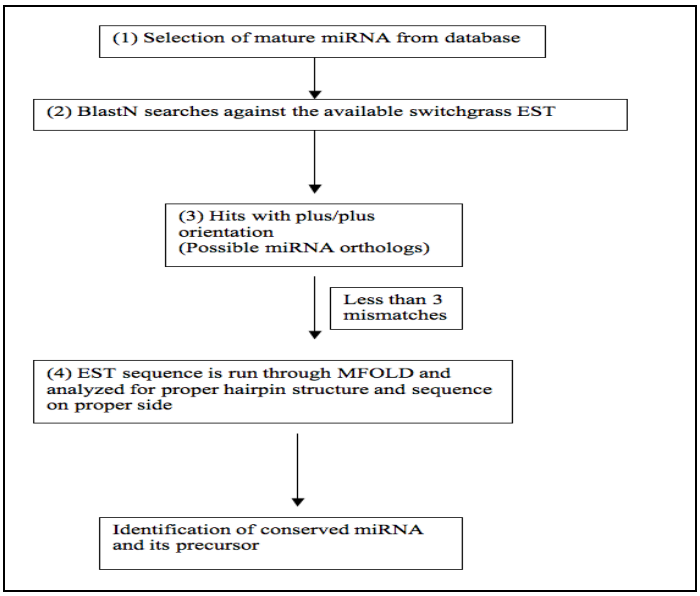


Figure 6. Basic outline of the computational strategy.

3.9: Sequence analysis of small RNA libraries

All sequences with perfect matches to both adaptor sequences were removed and the small RNAs between the adaptors were extracted. Small RNA sequences shorter than 17-nt and longer than 28-nt were discarded assuming that these fragments are not the Dicer products and might represent the degradation products from the larger RNAs. The presence of small, cloned RNAs as products of discrete miRNAs and siRNAs, as opposed to random RNA breakdown products can be determined bioinformatically. The switchgrass small RNAs as breakdown products from non-coding RNAs have been determined by using blast searches against databases such as the genomic tRNA database (http://gtrnadb.ucsc.edu/blast.html), and the European ribosomal RNA database (http://bioinformatics.psb.ugent.be/webtools/rRNA/), as the rRNAs and tRNAs are highly conserved among plant species. Conserved miRNAs were removed by blast searches

against the miRBase (http://www.mirbase.org/). In order to remove sequences for coding mRNAs, BLAST searches were performed against the plant EST database and any small RNAs with perfect matches in the plus/plus orientation were removed. All others were considered putative small RNAs (miRNA and siRNA). The remaining small RNA population has been carefully analyzed for the presence of novel miRNAs by searches against the switchgrass EST database, and ones that showed perfect matches were candidate small RNAs. Such ESTs were extracted and used for predicting fold-back structure using mFOLD (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi).

Secondary structure predictions

The novel mature miRNAs were then used as queries when searching the NCBI switchgrass EST database. The NCBI blastN program was used to find EST matches for ~90% of the new sequence (18 out of 20 for a 20nt novel miRNA). Hits among the ESTs with sense orientation (plus/plus orientation) with less than 3 mismatches were considered candidate precursors for the novel miRNAs. If multiple hits were found among the ESTs for a miRNA, those were extracted, aligned and a unique sequence was used for predicting the secondary structure to the miRNA precursor. The flanking region of 300bp upstream and downstream to the mature miRNA sequence was used for fold-back structure predictions using the mFOLD (http://mfold.bioinfo.rpi.edu/cgi-bin/rna-form1.cgi) program. The structures were then analyzed to make sure that the mature miRNA sequence was found on an arm of the hairpin structure.

3.10: Small RNA blot analysis

Twenty micrograms of LMW RNA were resolved on a denaturing 15% polyacrylamide/8M urea gel. The RNA was then electrophoretically transferred to a Hybond-N+ (Amersham) membrane, using a wet-blot transfer unit. The membrane was then UV cross-linked and baked at 80°C for 1 hour. DNA oligonucleotides, complementary to the miRNA sequences were end-labeled with γ-^{32}P-ATP using T4 polynucleotide kinase (Invitrogen), and used as probe. Blots were pre-hybridized for at least one hour and hybridized overnight using PerfectHYB+ buffer (Sigma) at 38°C. Blots were washed three times with 2xSSC/0.1%SDS, at 50°C. The blots were briefly air dried and exposed to a phosphoscreen. Images were acquired by scanning the phosphoscreen with a Typhoon scanner.

3.11: Target Predictions

The identified mature miRNA sequences were used for predicting RNA targets in switchgrass. The miRNA sequences were used as queries and searched against the ESTs from NCBI's switchgrass database for the complementary sequences. The following criteria were used for the target predictions: a total of four or fewer mismatches were allowed in the complementary region, however no mismatches were allowed in positions 10 and 11, which is the predicted cleavage site. Additionally, two continuous mismatches were not allowed. The annotation for the target EST has been determined by searching for homologous sequences in rice (http://rice.plantbiology.msu.edu/blast.shtml).

3.12: Validation of predicted miRNA targets using modified 5'-RACE assay

Modified 5'-RACE assay was performed using the GeneRacer Kit (Invitrogen, USA). Briefly a 5'-RNA adaptor was ligated to the RNA and subsequently a reverse transcription reaction was performed. The resulting cDNA was then used as a template for PCR amplification using primers (GeneRacer 5' primer and gene specific 3' primers). To ensure amplification of the desired product, a second, nested PCR was then carried out using nested primers (GeneRacer 5' nested primers and a gene-specific 3' nested primer). The amplified products were then gel purified, cloned into pGEM-T easy (Promega WI USA) and sequenced.

CHAPTER IV

RESULTS

4.1: Identification of conserved miRNAs using a computational approach

Conserved miRNAs can be predicted using a computational approach (Sunkar and Jagadeeswaran, 2008; Zhang et al., 2006), provided sufficient number of genomic or EST sequences are available for a plant species in question. For switchgrass, currently ~442,269 EST sequences are available at the NCBI database, while there are no genomic sequences available in the GSS (genomic survey sequence), HTGS (high-throughput genomic sequence), or the WGS (whole-genome shotgun reads) databases. A total of sixteen conserved miRNA families (miR156, miR159, miR160, miR164, miR166, miR167, miR169, miR171, miR319, miR394, miR397, miR399, miR408, miR437, miR444, and miR528) were identified in switchgrass using the computational approach (Table 3). Of these, thirteen families are conserved between monocots and dicots (miR156, miR159, miR160, miR164, miR166, miR167, miR169, miR171, miR319, miR394, miR397, miR399, and miR408), whereas miR437, miR444, and miR528 are conserved only in monocots. The predicted fold-back structures for these miRNA families share similar features with their counterparts found in *Arabidopsis* and rice (Figure 7).

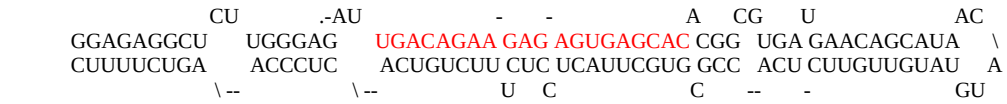
Because the number of ESTs available for switchgrass is limited, the computational analysis could not identify all conserved miRNA families in switchgrass. Furthermore, this approach is unlikely to identify novel species-specific miRNAs. In order to identify near complete set of conserved miRNAs as well as novel miRNAs in switchgrass, experimental approach was undertaken.

Table 3. Identified conserved miRNA homologs in switchgrass using a computational approach.

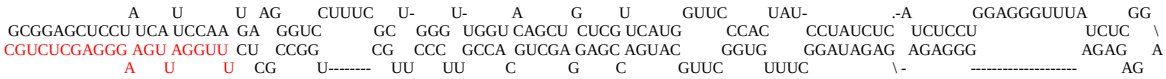
MicroRNA family	MicroRNA sequence
156	UGACAGAAGAGAGCGAGCAC
159	UUUGGAUUGAAGGGAGCUCUG
160	UGCCUGGCUCCCUGUAUGCCG
164	UGGAGAAGCAGGGCACGUGCA
166	UCGGACCAGGCUUCAUUCCCC
167	UGAAGCUGCCAGCAUGAUCUA
169	CAGCCAAGGAUGACUUGCCGA
171	UGAUUGAGCCGUGCCAAUAUC
319	UUGGACUGAAGGGUGCUCCC
394	UUGGCAUUCUGUCCACCUCC
397	UUAUUGAGUGCAGCGUUGAUG
399	UGCCAAAGGAGAUUUGCCCAG
408	CUGCACUGCCUCUUCCCUGG
437	UUCAGUUUGAAGAGAUUGAAA
444	UGCAGUUGCUGCCUCAAGCU
528	UGGAAGGGGCAUGCAGAGGAG

Figure 7. Predicted hairpin-like structures for the identified conserved miRNAs in switchgrass using the miRNA precursor sequences. The mature miRNA sequence is colored in red in each hairpin-like structure.

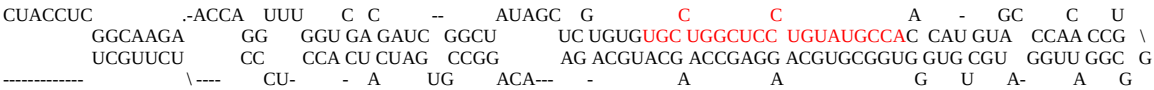
miR156



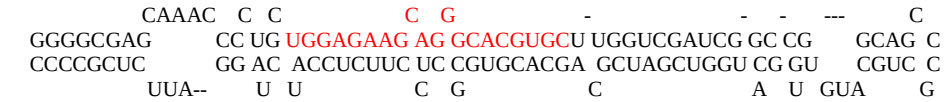
miR159



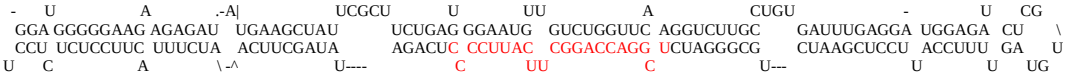
miR160



miR164

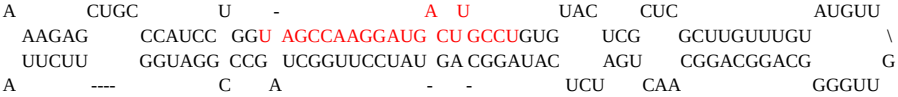


miR166

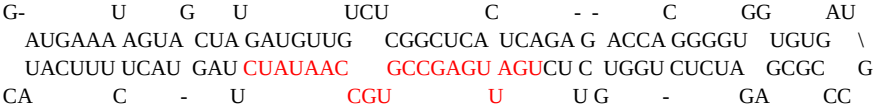


miR167

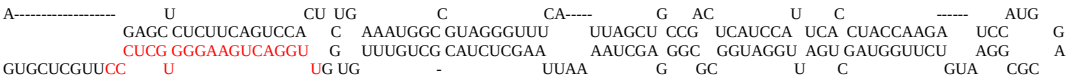
miR169



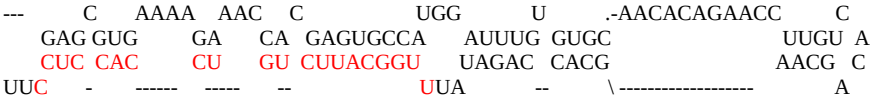
miR171



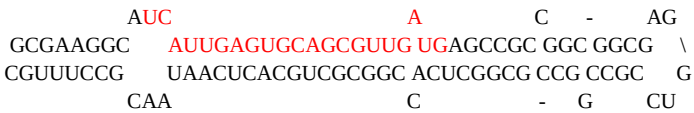
miR319



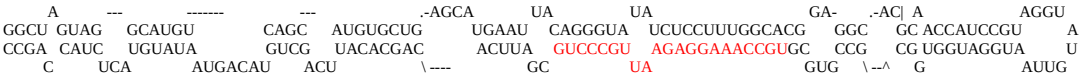
miR394



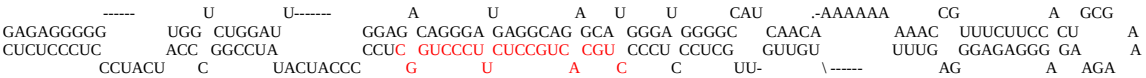
miR397



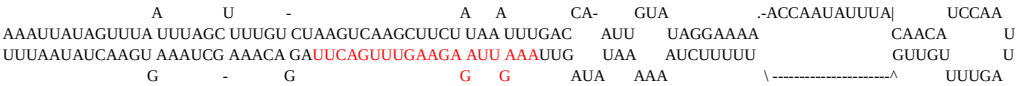
miR399



miR408

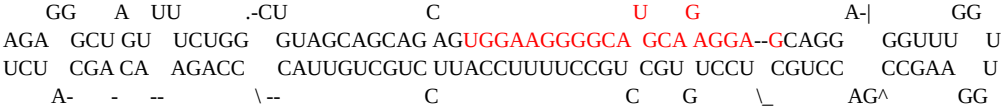


miR437



miR444

miR528



4.2: Identification of conserved miRNAs in switchgrass using an experimental approach.

Deep sequencing of small RNA libraries has been extremely successful in identification of not only conserved miRNAs but also novel miRNAs in several plant species including *Arabidopsis*, rice, *Populus*, and *Physcomitrella* (Sunkar and Zhu, 2004; Arazi et al., 2005; Sunkar et al., 2005b; Sunkar et al., 2005a; Talmor-Neiman et al., 2006a; Talmor-Neiman et al., 2006b; Fahlgren et al., 2007; Addo-Quaye et al., 2008; Lu et al., 2008; Sunkar et al., 2008; Zhu et al., 2008). In order to identify as many novel miRNAs as possible in switchgrass, three small RNA libraries were constructed from three diverse tissues, i.e., 3-week-old seedlings, inflorescence and emerging tillers and subjected to deep sequencing. Inflorescence and emerging tillers were chosen for constructing small RNA libraries for the following reasons. Previous reports suggested that small RNA population in plants is highly diverse in inflorescences (Lu et al., 2005). Previous work has shown that the greater number of tillers will contribute to high biomass production (Das et al., 2004) although the molecular basis of tillering is not well understood. In order to determine if miRNAs are playing a role in tillering, a small RNA profile was created from the emerging tillers.

Analysis of small RNA library generated from switchgrass seedlings

A small RNA library generated from switchgrass seedlings was subjected to pyrosequencing, which yielded 21,999 small RNA sequences ranging from 18 to 27 nt in size (Table 4). Figure 8 shows the distribution of sequences ranging in size from 18 to 26 nucleotides. After the removal of redundant sequences, 15,637 unique sequences were established. Of these, approximately 1,100 small RNAs appeared to be degraded products from protein-coding mRNAs, and another 2,084 sequences were mapped to

other noncoding RNA sequences, which were eliminated from further analysis. For the remaining unique small RNAs, homology searches against the miRBase resulted in identification of 34 conserved miRNAs belonging to 16 miRNA families (miR156, miR159, miR164, miR166, miR167, miR168, miR169, miR171, miR172, miR319, miR399, miR408, miR442, miR444, and miR528 in switchgrass (Table 5). Of the 21,999 raw sequences, only 269 small RNA sequences were identified as homologs of conserved miRNA families in switchgrass (Table 4). Thus, only a small fraction of the total small RNAs were miRNAs, and the remaining sequences are considered endogenous siRNAs, whose identity remains largely unknown.

Table 4. Summary of reads obtained from the seedlings small RNA library.

Total number of sequences	21,999
Unique sequences	15,637
Noncoding RNA/degraded mRNA	3,236
Conserved miRNA	272
New miRNA	4
Number of reads that could not be mapped to ESTs	~10,000

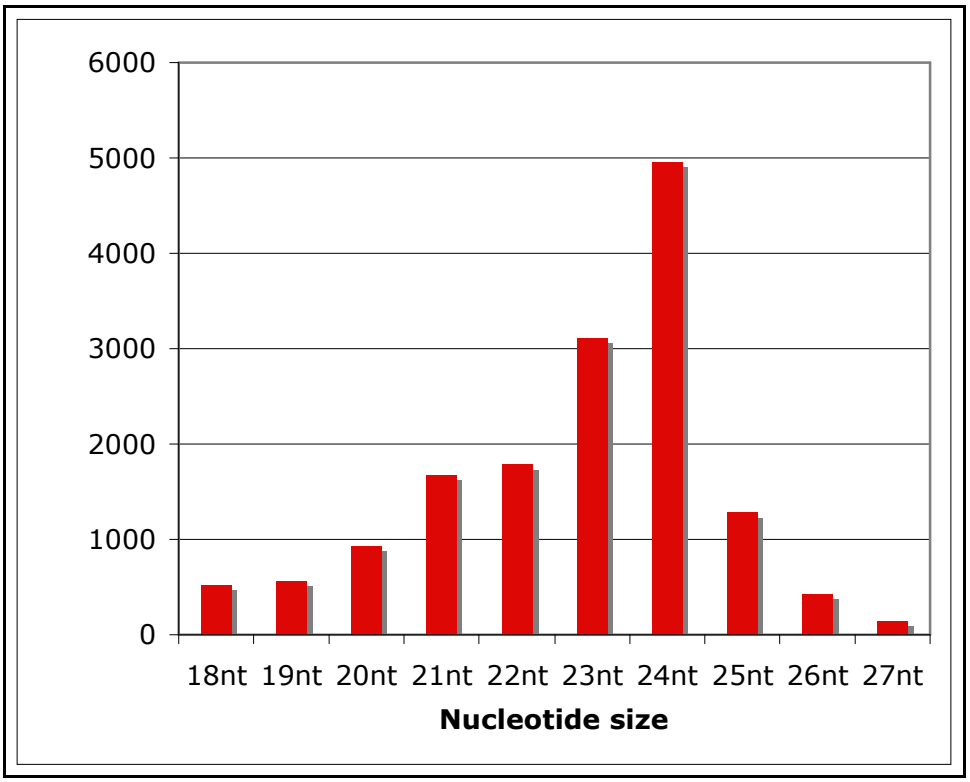


Figure 8. Size distribution and abundance of small RNAs in seedlings small RNA library.

Table 5. Identified conserved miRNAs and their frequency in seedlings small RNA library

miRNA family	miRNA Sequence (5'-3')	Frequency
miR156a,b	UGACAGAAGAGAGUGAGCAC	7
miR156e	UGACAGAAGAGAGCGAGCAC	23
miR156f	AGACAGAAGAGAGUGAGCAC	18
miR156k	UGACAGAAGAGAGAGAGCAC	2
miR159b	UUUGGAUUGAAGGGAGCUCUG	11
miR164a	UGGAGAAGCAGGGCACGUGCA	2
miR164c	UGGAGAAGCAGGGUACGUGCA	1
miR166	UCGGACCAGGCUUCAUUCCCC	10
miR167b	UGAAGCUGCCAGCAUGAUCUA	2
miR168	UCGCUUGGUGCAGAUCGGGAC	14
miR169a	CAGCCAAGGAUGACUUGCCGA	7
miR169c	CAGCCAAGGAUGACUUGCCGG	3
miR169d	CAGCCAAGGAUGACUUGCCUA	1
miR169b	UAGCCAAGGAUGACUUGCCGG	2
miR169k	UAGCCAAGGAUGACUUGCCUU	1
miR171g	UGAUUGAGCCGUGCCAAUAUC	8
miR172a	AGAAUCUUGAUGAUGCUGCAU	90
miR172c	ACUUGAUGAUGCUGCAGU	17
miR172b	GGAAUCUUGAUGAUGCUGCAU	24
miR172d	AGAAUCCUGAUGAUGCUGCAG	1
miR319	UUGGACUGAAGGGUGCUCCC	5
miR393	CUCCAAAGGGAUCGCAUUGAU	13
miR408	CUGCACUGCCUCUUCCCUGG	1
miR444	UGCAGUUGCUGCCUCAAGCU	1
miR528	UGGAAGGGGCAUGCAGAGGAG	5
Tas3-siRNA	UUGGGAGGAUUGAUAGGCGCUA	3

Within this library, the frequency of different miRNA family members was highly varied, as some miRNAs (miR172a) appeared as many as 90 times, whereas some others were found only once in the library. Of the 272 miRNA homologs found in the library, miR172 family was the most abundantly represented family with 132 reads. Within the miR172 family, miR172a alone appeared 90 times, and thus was the most abundantly

expressed member of the miR172 family. miR156 was the second most abundantly recovered miRNA family with a total number of 53 reads in the library. The frequencies of the remaining conserved miRNA were relatively low, i.e. miR168 family was represented by 14 reads, miR393 family by 13 reads, and miR159 family by 11 reads in the library. Two miRNA families (miR408, and miR444) appeared only once in the library (Table 5).

In plants, trans-acting siRNAs (TAS3-siRNAs) are highly conserved and their biogenesis is dependent on miR390, which is a conserved miRNA (Allen, *et al*. 2005). While, miR390 was not found in the analyzed small RNA library, a family member of the TAS3 siRNA family (TAS3b) was identified (Table 5), suggesting that TAS3 siRNAs are also conserved, and expressed in switchgrass.

While the first small RNA library provided a glimpse of conserved miRNA families expressed in switchgrass, it did not reveal all conserved miRNA families. For instance, miR160, miR162, miR390, miR394, miR395, miR396, miR397, miR398 and miR399, which are conserved were not found in this library. This could be due to the low-sequencing depth (about 22,000 reads) and suggested that sequencing to a greater depth than this is necessary to identify all conserved miRNAs. These results also pointed out that identification of novel miRNAs in switchgrass require sequencing small RNA libraries to even greater depth because novel miRNAs are expressed at extremely low abundance. . In order to identify additional conserved miRNA homologs as well as novel miRNAs in switchgrass, two other small RNA libraries were generated from inflorescence and emerging tillers and sequenced using sequencing-by-synthesis technology, which yielded several million (10-11 million) reads from each library.

Analysis of small RNA library generated from inflorescence and emerging tillers

The overall summary of the small RNAs obtained from sequencing small RNA populations from the inflorescence and emerging tillers is shown in Table 6.

Table 6. Summary of sequence analysis of small RNA libraries from inflorescence and emerging tillers.

	Inflorescence		Emerging tillers	
Small RNA category	Total reads	unique reads	Total reads	unique reads
tRNA, rRNA, snRNA etc.	2153056	199430	1518579	112572
Repeats	1172189	187883	1886253	184354
conserved miRNAs	720393	236	189054	230
pre-miRNAs	1149079	24501	403542	16670
EST	1606132	157529	1599648	132424
Could not be mapped to ESTs	8784958	4830653	7904002	3502843
Total	11746013	5175081	10658965	3757827

Deep sequencing of small RNA reads of 18 to 27-nt in size from the inflorescence and emerging tillers yielded 11,746,013 and 10,658,965 raw reads, respectively (Table 6 and Figure 9). Size-based analysis of small RNA reads revealed two peaks; one at the 21 nt and the other at 24 nt size (Figure 9). This observation is consistent with the small RNA populations in plants (Lu et al., 2005; 2006; Sunkar et al., 2005b; Fahlgren et al., 2007; Jagadeeswaran et al., 2009). In general, the majority of small RNAs corresponding to 21-nt size are likely miRNAs, whereas the 24-nt small RNAs correspond to heterochromatic siRNAs in plants (Lu et al., 2006; Sunkar and Zhu, 2007). The emerging tillers library was represented by slightly greater number of 24 nt small RNAs than the inflorescence library, while the opposite was true for 21-nt small RNAs in inflorescence library, i.e., more 21-nt small RNAs in inflorescence library compared to the emerging tillers library (Figure 9). After the removal of redundant sequences, 5,175,081 and

3,757,827 unique reads were obtained for inflorescence and emerging tillers, respectively (Table 6). The degraded products from the rRNAs, tRNAs as well as from the protein-coding mRNAs were removed from the analysis. For the remaining unique small RNAs, homology searches against the miRBase resulted in identification of all known miRNAs (some are highly conserved and some are known to be conserved in closely related plant species) belonging to 36 miRNA families in switchgrass (Table 7). A large proportion of the small RNA reads from both the libraries could not be mapped to any of the small RNA categories (Table 6). Most of these might represent endogenous siRNAs and degradation products from mRNAs in switchgrass although it is possible that some of these could be genuine miRNAs. However, in the absence of miRNA* sequence and the non-availability of the switchgrass genome for mapping and predicting fold-back structure for the miRNA precursor it is not possible to annotate them as novel miRNAs in switchgrass.

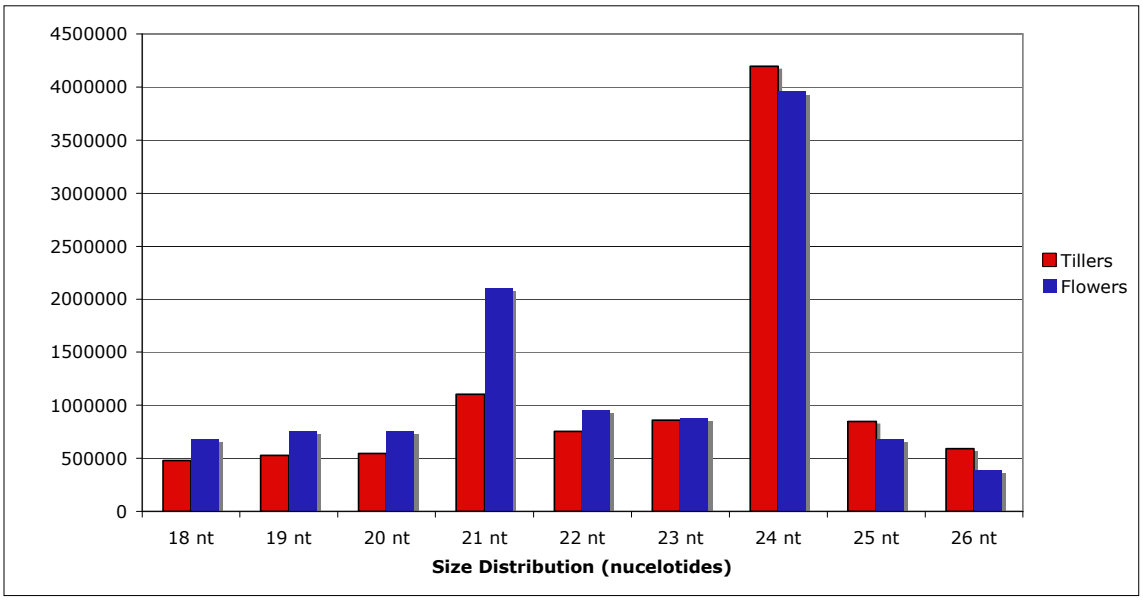


Figure 9: Size distribution and abundance of small RNAs in inflorescence and emerging tillers small RNA libraries.

The frequency of known miRNA families identified in switchgrass inflorescence and emerging tillers is shown in Table 8. In plants, most conserved miRNAs exists as multiple members transcribed from different loci, which are termed as miRNA families. The sequence analyses resulted in identification of 260 distinct miRNAs belonging to 36 miRNA families (miR156, miR159, miR160, miR162, miR164, miR165/166, miR167, miR168, miR169, mIR170/171, miR172, miR319, miR390, miR393, miR394, miR395; miR396; miR397; miR398, miR399, miR408, miR444, miR528, miR529, miR535, miR827, miR845, miR894, miR1318, miR1432, miR2118, miR2275, miR2910, miR2914, miR2915 and miR2916) in switchgrass. miR395 had almost an equal representation in each library (89 times in the inflorescence and 74 times in the emerging tiller), while miR399 had a higher representation in the inflorescences (84 times) in comparison to the emerging tillers (2 times), suggesting differential expression of miR399 in different tissues of switchgrass.

miR444 was present 429 times in the inflorescence library and 36 times in the emerging tillers library; while miR528 was present 96 times in the inflorescence library and 384 times in the emerging tillers library (Table 7). Of the monocot specific miRNA families, miR528 had the highest abundance in the emerging tillers. Other semi-conserved miRNA families identified in the deep-sequenced libraries included miR827, and miR2118. These families are expressed in some monocot species and some dicot species (Lu, et al. 2008; Fahlgren, et al. 2007; Jagadeeswaran, et al. 2009a; Arenas-Huertero, et al. 2009; Johnson, et al. 2009), but not in all monocots and dicots.

The overall abundance of miRNA family varied between the two libraries. miR165/166 family had the highest number of normalized reads (TPM, transcripts per

million), which is followed by miR168, miR156 and miR894 in the inflorescence small RNA library (Table 7). The same was also true for the small RNA populations in emerging tillers with the exception that the 4th most abundantly expressed miRNA family is miR528 (Table 7).

On the basis of number of miRNA family members recovered, miR170/171 family is the largest family as is represented by 18 members in switchgrass (Table 7). This is followed by miR165/166 (17 members), miR169 (16 members), miR159 and miR399 (13 members), miR172 (12 members), miR167 (11 members), miR156 (10 members), miR396 (8 members), each of the miR164 and miR395 (7 members), whereas several miRNA families such as miR394, miR398, miR827, miR845, miR894, miR1318, miR2275, miR2910, miR2914, miR2915 and miR2916 are represented by one member only in switchgrass (Table 7).

Within a miRNA family, the expression abundance of different loci appears to differ, which could confer a tissue- or cell-specific expression of different members. Therefore, it is important to assess which locus is highly expressed. The expression from different loci can be assessed from the frequency of their appearance in the library provided at least these members vary in one nucleotide. Interestingly, a greater disparity exists among different members of the same miRNA families, i.e., few variants/loci are most abundantly expressed than the others. Most distinct one is miR168, which appears to have 3 loci but only one of them is abundantly expressed (10023 TPM and 3716 TPM in inflorescence and tillers libraries, respectively) where as the other two loci are represented by 1-3 TPM in both the libraries. Similarly, some of the loci belonging to

miR169 are expressed as abundantly as about 60,000 TPM where as some others are expressed at very low levels (44 TPM).

Table 7. Frequency of conserved miRNA families and TAS3-siRNA found in the inflorescence and emerging tillers small RNA libraries. (TPM; Transcripts per million).

miRNA	miRNA sequence	Number of raw reads in the inflorescence library	Normalized reads in the inflorescence library (TPM)	Number of raw reads in the emerging tillers library	Normalized reads in the emerging tillers library (TPM)
miR156	CUGACAGAAGAUAGAGAGCAC	2	0	3	0
miR156	UUGACAGAAGAAAGAGAGCAC	1	0	0	0
miR156	UGACAGAAGAGAGCGAGCAC	14795	1258	4878	457
miR156	UGACAGAGGAGAGUGAGCAC	11	1	9	1
miR156	UGACAGAAGAGAGAGAGCAU	75	6	237	22
miR156	UGUCAGAAGAGAGUGAGCAC	18	2	19	2
miR156	UGACAGAAGAGAGUGAGCAC	20097	1708	16915	1585
miR156	UGACAGAAGAGAGUGAGCACA	19887	1691	16718	1567
miR156	UUGACAGAAGAUAGAGAGCAC	2	0	3	0
miR156	CGACAGAAGAGAGUGAGCAC	19933	1694	16722	1567
miR156	UGACAGAAGAAAGAGAGCAC	1	0	0	0
miR156	UGACAGAAGAUAGAGAGCAC	2	0	3	0
miR156	UGACAGAAGAGAGGGAGCAC	17	1	8	1
miR156	UGACAGAAGAGAGAGAGCACA	74	6	238	22
miR156	CGACAGAAGAGAGUGAGCAUA	960	82	505	47
miR158	UCCCAAAUGUAGACAAAGCA	1	0	3	0
miR158	CCCCAAAUGUAGACAAAGCA	0	0	3	0
miR159	UUUGGACUGAAGGGAGCUCUA	14	1	2	0
miR159	UUGGAUUGAAGAGAGCUCCC	1	0	1	0
miR159	UUUGGAUUGAAAGGAGCUCUU	21	2	8	1
miR159	UUUGGAUUGAAGGGAGCUCUU	9951	846	4248	398
miR159	UUUGGAUUGAAGGGAGCUCUG	10553	897	4815	451
miR159	UUUGGAUUGAAGGGAGCUCUA	9954	846	4256	399
miR159	CUUGGAGUGAAGGGAGCUCUC	1	0	0	0
miR159	UUUGGAUUGAAGGGAGCUCCU	9911	843	4203	394
miR159	UUGGAUUGAAGGGAGCUCCA	3666	312	2028	190
miR159	AUUGGAUUGAAGGGAGCUCCU	143	12	52	5
miR159	AUUGGAUUGAAGGGAGCUCCG	143	12	52	5
miR159	AUUGGAUUGAAGGGAGCUCCA	143	12	52	5
miR159	CUUGGAUUGAAGGGAGCUCUA	2802	238	1582	148
miR159	CUUGGAUUGAAGGGAGCUCCU	118	10	37	3
miR159	CUUGGAUUGAAGGGAGCUCCC	119	10	37	3
miR160	UGCCUGGCUCCCUGAAUGCCA	1	0	0	0
miR160	UGCCUGGCUCCCUGUAUGCCG	1955	166	227	21

miR160	UGCCUGGCUCCCUGUAUGCCA	2041	174	240	22
miR160	UGCCUGGCUCCUUGUAUGCCA	2	0	0	0
miR160	CGCCUGGCUCCCUGCAUGCCG	3	0	0	0
miR160	CGCCUGGCUCCCUGCAUGCCA	3	0	0	0
miR160	CGCCUGGCUCCCUGUAUGCCA	1921	163	226	21
miR160	CGCCUGGCUCCUUGUAUGCCA	2	0	0	0
miR160	UGCCUGGCUCCCUGGAUGCCA	0	0	4	0
miR160	UGCCUGGCUCCCUGUAUGCC	2042	174	239	22
miR160	UGCCUGGCUCCCUGCAUGCCA	3	0	0	0
miR162	UCGAUAAACCUCUGCAUCCA	272	23	76	7
miR162	UCGAUAAACCUCUGCAUCCAG	271	23	74	7
miR164	UGGAGAAGCAGGGCACAUGCU	14	1	7	1
miR164	UGGAGAAGCAGGGCACGUGCU	8195	697	5297	496
miR164	UGGAGAAGCAGGGCACGUGCG	8175	695	5290	496
miR164	UGGAGAAGCAGGGCACGUGCA	8248	701	5351	502
miR164	UGGAGAAGCAGGGCACGUGAG	7938	675	5195	487
miR164	UGGAGAAGCAGGGUACGUGCA	172	15	40	4
miR164	UGGAGAAGCAGGACACGUGAG	1205	102	101	9
miR165	UCGGACCAGGCUUCAUCCCCC	1433	122	220	21
miR166	UCGGACCAGGCUUCAUUCCCCU	704855	59918	156963	14711
miR166	UCGGACCAGGCUUCAUUCCCCC	704444	59883	156936	14709
miR166	UCGGACCAGGCUUCAUUCCUU	708773	60251	163110	15287
miR166	UCGGACCAGGCUUCAUUCCUG	708669	60242	163095	15286
miR166	UCGGACCAGGCUUCAUUCCUC	709831	60341	163227	15298
miR166	UCGGACCAGGCUUCAUUCCUA	708680	60243	163097	15286
miR166	CCGGACCAGGCUUCAUCCCAG	4	0	1	0
miR166	UCGGGCCAGGCUUCAUCCCCC	1	0	0	0
miR166	UCGGACCAGGCUUCAUUCCCU	757637	64405	166234	15580
miR166	UCGGACCAGGCUUCAUUCCCC	772253	65648	168507	15793
miR166	UCGGACCAGGCUCCAUUCCUU	652	55	108	10
miR166	UCGGACCAGGCUUCAUUCCU	753874	64085	165191	15482
miR166	UCGGACCAGGCUUCAUUCCC	778616	66189	169537	15890
miR166	UCCGGACCAGGCUUCAUUCCC	692378	58858	154541	14484
miR166	UCGGACCAGGCUUCAAUCCCU	15572	1324	2523	236
miR166	UCGAACCAGGCUUCAUUCCCC	691	59	123	12
miR166	UCGGAUCAGGCUUCAUUCCUC	1244	106	206	19
miR166	UCGGACCAGGCAUCAUUCCUU	523	44	32	3
miR167	AUCAUGCAUGACAGCCUCAUUU	5	0	5	0
miR167	UGAAGCUGCCAGCAUGAUCUU	9985	849	7581	711
miR167	UGAAGCUGCCAGCAUGAUCUG	10233	870	7899	740
miR167	UGAAGCUGCCAGCAUGAUCUC	9971	848	7580	710
miR167	UGAAGCUGCCAGCAUGAUCUA	9990	849	7603	713
miR167	UCAAGCUGCCAGCAUGAUCUA	1094	93	592	55
miR167	UGAAGCUGCCAGCAUGAUCUGG	9995	850	7584	711
miR167	UGAAGCUGCCAGCAUGAUCUAA	9864	839	7447	698
miR167	UAAGCUGCCAGCAUGAUCUUG	169	14	39	4
miR167	GGAAGCUGCCAGCAUGAUCCU	63	5	57	5

miR167	UGAAGCUGCCAACAUGAUCUG	13	1	6	1
miR168	UCGCUUGGGCAGAUCGGGAC	22	2	4	0
miR168	UCGCUUGGUGCAGAUCGGGAC	117910	10023	39653	3716
miR168	UCGCUUGGUGCAGGUCGGGAA	40	3	18	2
miR169	CAGCCAAGGAUGACUUGCCGG	460	39	548	51
miR169	CAGCCAAGGAUGACUUGCCGA	454	39	544	51
miR169	UAGCCAAGGAUGACUUGCCUGC	170	14	197	18
miR169	UAGCCAAGGAUGACUUGCCUG	185	16	233	22
miR169	UAGCCAAGGAUGACUUGCCUA	189	16	236	22
miR169	UAGCCAAGGAUGACUUGCCGG	611	52	726	68
miR169	UAGCCAAGGAUGACUUGCCCA	177	15	227	21
miR169	UGAGCCAAGGAUGACUUGCCG	444	38	531	50
miR169	UCAGCCAAGGAUGACUUGCCG	455	39	542	51
miR169	UAGCCAAGAAUGACUUGCCUA	49	4	56	5
miR169	UGAGCCAAGGAUGGCUUGCCG	3	0	0	0
miR169	UGAGCCAAAGAUGACUUGCCG	1	0	1	0
miR169	AAGCCAAGGAUGACUUGCCUG	176	15	197	18
miR169	AAGCCAAGGAUGACUUGCCUA	177	15	196	18
miR169	AAGCCAAGGAUGACUUGCCGG	448	38	535	50
miR169	AAGCCAAGGAUGACUUGCCGA	446	38	532	50
miR169	UAGCCAAGGAUGACUUGCUCG	1	0	10	1
miR169	GAGCCAAGGAUGACUUGCCGU	444	38	531	50
miR169	GAGCCAAGGAUGACUUGCCGG	447	38	535	50
miR169	GAGCCAAGGAUGACUUGCCGC	444	38	531	50
miR169	GGAGCCAAGGAUGACUUGCCG	444	38	531	50
miR169	GAGCCAAGAAUGACUUGCCGG	1	0	1	0
miR169	UAGCCAAGGACGACUUGCCUG	1	0	0	0
miR169	UAGCCAAGGACGACUUGCCUA	1	0	1	0
miR169	UAGCCAAGGAUGAAUUGCCGG	3	0	1	0
miR169	GAGCCAAAGAUGACUUGCCGG	1	0	1	0
miR169	AAGCCAAGGAUGAAUUGCCGG	3	0	1	0
miR170	UGAUUGAGCCGUGUCAAUAUC	4	0	6	1
miR171	UGAGCCGUGCCAAUAUCACAU	4	0	3	0
miR171	UUGAGCCGUGCCAAUAUCAUG	7	1	6	1
miR171	UUGAGCCGUGCCAAUAUCAUC	7	1	7	1
miR171	UUGAGCCGUGCCAAUAUCACU	9	1	5	0
miR171	UUGAGCCGUGCCAAUAUCACG	10	1	7	1
miR171	UUGAGCCGUGCCAAUAUCACA	8	1	5	0
miR171	UGACUGAGCCGUGCCAAUAUC	3	0	0	0
miR171	UGAUUGAGCCGCGCCAAUAUCU	11	1	2	0
miR171	UGAUUGAGCCGCGCCAAUAUC	11	1	2	0
miR171	UGAUUGAGCCGUGCCAAUAUU	1456	124	1838	172
miR171	UGAUUGAGCCGUGCCAAUAUC	1475	125	1868	175
miR171	GGAUUGAGCCGCGUCAAUAUC	6	1	2	0
miR171	UGAUUGAGCCGCGCCAAUAU	11	1	2	0
miR171	AGAUUGAGCCGCGCCAAUAUC	10	1	1	0
miR171	UUGAGCCGUGCCAAUAUCAC	1378	117	1726	162

miR171	UGAUUGAGUCGUGCCAAUAUC	3	0	5	0
miR171	GGAUUGAGCCGCGCCAAUAUC	10	1	1	0
miR171	UGAUUGAGCCGCGUCAAUAUC	6	1	0	0
miR171	UAAUUGAGCCGUGCCAAUAUC	1379	117	1734	163
miR171	UGAGCCGCGCCAAUAUCACAU	0	0	1	0
miR171	UUGAGCCGAACCAAUAUCACC	4	0	1	0
miR171	UUGAGCCGCGCCAAUAUCAC	11	1	3	0
miR171	AUGAGCCGAACCAAUAUCACU	4	0	1	0
miR171	UUGAGCCGCGCCAAUAUCACU	0	0	1	0
miR171	UUGAGCCGCGCCAAUAUCACA	0	0	1	0
miR171	GUGAGCCGAACCAAUAUCACU	4	0	1	0
miR171	UGAGCCGUGCCAAUAUCACGA	6	1	5	0
miR172	AGAAUCUUGAUGAUGCUGCAU	9042	769	1943	182
miR172	AGAAUCUUGAUGAUGCUGCAG	8951	761	1919	180
miR172	GGAAUCUUGAUGAUGCUGCA	9003	765	1913	179
miR172	UGAAUCUUGAUGAUGCUACAU	18	2	2	0
miR172	UGAAUCUUGAUGAUGCUACAC	18	2	2	0
miR172	GGAAUCUUGAUGAUGCUGCAGCAG	41	3	5	0
miR172	GGAAUCUUGAUGAUGCUGCAU	9035	768	1921	180
miR172	GGAAUCUUGAUGAUGCUGCAG	8949	761	1908	179
miR172	AGAAUCCUGAUGAUGCUGCAG	17	1	5	0
miR172	AGAAUCCUGAUGAUGCUGCAA	17	1	5	0
miR172	UGAAUCUUGAUGAUGCUGCAC	8922	758	1909	179
miR172	AGAAUCUUGAUGAUGCUGCA	9044	769	1939	182
miR319	CUUGGACUGAAGGGAGCUCCC	1	0	0	0
miR319	UUUGGACUGAAGGGAGCUCCU	15	1	2	0
miR319	CUUGGACUGAAGGGAGCUCC	5	0	2	0
miR319	UUGGACUGAAGGGUGCUCCC	1800	153	169	16
miR319	AUUGGACUGAAGGGAGCUCCC	1	0	0	0
miR319	UUGGACUGAAGGGAGCUCC	14	1	2	0
miR319	UUGGACUGAAGGGAGCUCCUU	1	0	0	0
miR319	UUGGACUGAAGGGAGCUCCCU	1	0	0	0
miR319	UUGGACUGAAAGGAGCUCCU	0	0	1	0
miR319	UUGGACUGAAGGGAGCUCCU	5	0	2	0
miR319	UUGGACUGAAGGGAGCUCCC	5	0	2	0
miR390	AAGCUCAGGAGGGAUAGCGCC	81	7	16	1
miR390	GAGCUCAGGAGGGAUAGCGCC	78	7	14	1
miR390	AAGCUCAGGAGGGAUAGCACC	0	0	2	0
miR393	UCCAAAGGGAUCGCAUUGAUC	161	14	152	14
miR393	UCCAAAGGGAUCGCAUUGAUCU	129	11	128	12
miR393	UCCAAAGGGAUCGCAUUGAUCC	130	11	129	12
miR394	UUGGCAUUCUGUCCACCUCC	528	45	45	4
miR395	CUGAAGUGUUUGGGGGAACUCC	263	22	184	17
miR395	CUGAAGUGUUUGGGGGAACUC	264	22	187	18
miR395	UUGAAGUGUUUGGGGGAACUC	264	22	187	18
miR395	CUGAAGUGUUUGGGGGGACUC	1	0	0	0
miR395	GUGAAGUGCUUGGGGGAACUC	14	1	29	3

miR395	AUGAAGUGUUUGGGGGAACUU	262	22	184	17
miR395	AUGAAGUGUUUGGGGGAACUC	264	22	187	18
miR395	GUGAAGUGUUUGGGGGAACUC	264	22	188	18
miR396	UUCCACAGGCUUUCUUGAACUG	1788	152	305	29
miR396	CUCCACAGGCUUUCUUGAACUG	1784	152	307	29
miR396	UUCCACAGCUUUCUUGAACUU	3279	279	425	40
miR396	UUCCACAGCUUUCUUGAACUG	3307	281	434	41
miR396	UUCCACAGCUUUCUUGAACUA	3271	278	423	40
miR396	UCCACAGGCUUUCUUGAACUG	1827	155	318	30
miR396	UUCCACGGCUUUCUUGAACC	0	0	1	0
miR396	UCUCCACAGGCUUUCUUGAACU	1712	146	301	28
miR396	UUCCACGGCUUUCUUGAACUU	0	0	1	0
miR396	UUCCACGGCUUUCUUGAACUG	0	0	1	0
miR396	UUCCACAGCUUUCUUGAACU	3369	286	443	42
miR397	CCAUUGAGUGCAGCGUUGAUG	25	2	34	3
miR397	UUAUUGAGUGCAGCGUUGAUG	25	2	34	3
miR397	UCAUUGAGUGCAGCGUUGAUG	25	2	34	3
miR397	UCAUUGAGUGCAGCGUUGAUGU	1	0	3	0
miR397	AUUGAGUGCAGCGUUGAUGA	323	27	451	42
miR398	UGUGUUCUCAGGUCGCCCCUG	11	1	14	1
miR398	UGUGUUCUCAGGUCACCCCUU	1	0	0	0
miR398	UGUGUUCUCAGGUCACCCCUG	1	0	0	0
miR399	UGCCAAAGGAGAGUUGCCCUG	109	9	7	1
miR399	UGCCAAAGGAGAGUUGCCCUA	109	9	6	1
miR399	UGCCAAAGGAGAGCUGCCCUG	42	4	1	0
miR399	UGCCAAAGGAGAGCUGCCCUA	39	3	1	0
miR399	UGCCAAAGGAAAUUUGCCCCG	1	0	0	0
miR399	CGCCAAAGGAGAGUUGCCCUG	62	5	1	0
miR399	CGCCAAAGGAGAGUUGCCCUC	62	5	1	0
miR399	UGCCAAAGGAGAUUUGCUCGU	1	0	0	0
miR399	UGCCAAAGGAGAUUUGCCCUG	82	7	1	0
miR399	UGCCAAAGGAGAUUUGCUCAC	1	0	0	0
miR399	UGCCAAAGGAGAUUUGCCCGG	82	7	2	0
miR399	UGCCAAAGGAGAUUUGCCCCU	82	7	1	0
miR399	UGCCAAAGGAGAUUUGCCCCG	82	7	1	0
miR399	UGCCAAAGGAGAAUUGCCC	83	7	3	0
miR399	UGCCAAAGGAGAUUUGCCCAG	84	7	1	0
miR399	UGCCAAAGAAGAUUUGCCCUG	1	0	0	0
miR399	UGCCAAAGAAGAUUUGCCCCG	1	0	0	0
miR399	UGCCAAAGGAGAAUUGCCCUG	77	7	2	0
miR399	UGCCAAAGGAGAUUUGCCUCG	1	0	0	0
miR408	AUGCACUGCCUCUUCCCUGGC	170	14	114	11
miR408	CUGCACUGACUCUUCCCUGGC	1	0	0	0
miR408	UGCACUGCCUCUUCCCUGGCU	173	15	122	11
miR408	AUGCACUGCCUCUUCCCUGG	164	14	114	11
miR408	CUGCACUGCCUCUUCCCUGGC	177	15	119	11
miR408	UGCACUGCCUCUUCCCUGGCUG	171	15	112	10

miR444	UGCAGUUGCUGCCUCAAGCUU	368	31	76	7
miR444	UGUUGUCUCAAGCUUGCUGCC	2	0	1	0
miR444	UGCAGUUGUUGCCUCAAGCUU	1086	92	70	7
miR444	UUGUGGCUUUCUUGCAAGUUG	101	9	9	1
miR444	UGCAGUUGUUGUCUCAAGCUU	3495	297	227	21
miR528	UGGAAGGGGCAUGCAGAGGAG	1126	96	4095	384
miR529	AGAAGAGAGAGAGCACAGCCC	2	0	0	0
miR529	CUGUACCCUCUCUCUUCUUC	5	0	19	2
miR529	AGAAGAGAGAGAGUACAGCUU	629	53	285	27
miR529	AGAAGAGAGAGAGUACAGCCC	546	46	295	28
miR535	UGACAACGAGAGAGAGCACGC	1490	127	372	35
miR535	UGACAACGAGAGAGAGCACGCU	1480	126	364	34
miR535	UGACAACGAGAGAGAGCACGCG	1480	126	363	34
miR827	UUAGAUGACCAUCAACAAACU	6	1	2	0
miR827	UUAGAUGACCAUCAGCAAACA	9797	833	2062	193
miR829	AGCUCUGAUACCAAAUGAUGGAAU	2	0	0	0
miR845	UAGCUCUGAUACCAAUUGAUA	2	0	0	0
miR845	UGGCUCUGAUACCAAUUGAUG	2	0	0	0
miR845	AGGCUCUGAUACCAAUUGAUG	7	1	0	0
miR845	CGGCUCUGAUACCAAUUGAUG	3	0	1	0
miR858	UUUCGUUGUCUGUUCGACCUU	1	0	0	0
miR894	CGUUUCACGUCGGGUUCACC	18368	1561	5267	494
miR1050	UGACCACCUUGAUUCCGGCCU	1	0	0	0
miR1171	UGGAGUGGAGUGGAGUGGAGUGG	2	0	0	0
miR1318	UCAGGAGAGAUGACACCGAC	23	2	23	2
miR1432	CUCAGGAGAGAUGACACCGAC	22	2	22	2
miR1432	AUCAGGAGAGAUGACACCGAC	13	1	15	1
miR2086	GACAUGAAUGCAGAACUGGAA	0	0	1	0
miR2112	CUUUAUAUCCGCAUUUGCGCA	1	0	0	0
miR2118	UUCCUGAUGCCUCCCAUGCCUA	11	1	7	1
miR2118	UUCCCGAUGCCUCCCAUUCCUA	32	3	18	2
miR2118	UUCCUGAUGCCUCCUAUUCCUA	2	0	0	0
miR2118	UUCCUGAUGCCUCUCAUUCCUA	2	0	1	0
miR2275	UUCAGUUUCCUCUAAUAUCUCA	1	0	6	1
miR2910	UAGUUGGUGGAGCGAUUUGUC	2007	171	3468	325
miR2914	CAUGGUGGUGACGGGUGACGGAG	472	40	564	53
miR2915	CCCGUCUAGCUCAGUUGGUA	22	2	22	2
miR2916	UGGGGACUCGAAGACGAUCAUAU	7	1	12	1
TAS3a (siRNA)	UCUUGACCUUGUAAGACCCAA	173	15	26	2

4.3. Identification of novel miRNAs in switchgrass

Identifying novel miRNAs that are conserved in closely related species or are species-specific is a difficult task given the fact that only a minor proportion of plant

small RNA population is miRNAs, whereas the vast majority of them are endogenous siRNAs. Because of this challenge, the plant small RNA community has provided few guidelines for annotation purposes (Meyers et al., 2008). As per their norms, a small RNA can be annotated as "novel miRNA" provided that miRNA* sequence corresponding to the novel small RNA appears in the small RNA library (Meyers et al., 2008). MicroRNA* sequences are relatively less abundant than are their miRNA counterparts. Our sequence analysis revealed 13 small RNAs as novel miRNAs based on sequencing of miRNA* reads in the small RNA libraries (Table 8). In the absence of miRNA* sequence, conservation of the miRNA sequence in related plant species coupled with the predictable fold-back structure for the precursor sequences could partially satisfy the classification of an siRNA as "novel miRNA" (Jagadeeswaran et al., 2009). We analyzed if any of the newly identified unique small RNAs are conserved in closely related monocots using BLAST searches against the NCBI EST database. Surprisingly, 6 of the novel small RNAs (7724135, 6651927, 5564248, 8018588, 4001019 and 185087) are conserved at least in one other monocot (maize, sorghum, sugarcane, rice, *Cenchrus ciliaris*), (Figure 10). Fold-back structures for the novel miRNA precursors from maize, sorghum, sugarcane, rice, *Cenchrus ciliaris* could be predicted using their precursor sequences (Figure 10). Interestingly, miRNA* reads were also recovered for four (7724135, 5564248, 4001019 and 185087) of the six conserved miRNAs. On the basis of appearance of miRNA* read (13 small RNAs of which four are also conserved) in the library and their conservation (a total of 6 are conserved but 4 of them are included in the other group because of the miRNA* reads were recovered) in related monocots, 15 small RNAs were annotated as "novel miRNAs" in switchgrass. Taken together, six novel

miRNAs can be annotated as monocot-specific miRNAs and homologs for the remaining 9 novel miRNAs could not be found in other monocots, thus could be annotated as switchgrass-specific miRNAs for now. Interestingly, the frequency of a few novel miRNAs (s3496977, 7724135 and 850747) is substantially higher and even greater than the frequency of several highly conserved miRNAs such as miR396, miR393, miR398 etc.

Table 8: Identification of novel miRNAs in switchgrass based on the appearance of miRNA* sequences or conservation in related monocots. (Bold letters indicate the conserved miRNAs in other related monocots)**.**

miRNA	**miRNA sequence**	**Frequency in inflorescence small RNA library**	**Frequency in emerging tillers small RNA library**	**miRNA* sequence**	**Frequency in inflorescence small RNA library**	**Frequency in emerging tillers small RNA library**
8008250	UUCAGGACCGGCUUCACACGUGAA	212	36	CACGUGUGAAGCCGGUCCUGAAGC	82	32
6815382	UAAUGACGGUAAUUAAUUGAUGAU	319	87	UCACGCGGUCAAAGGCCUCAUUAG	41	22
3496977	AUUUCACGGAGUUGUGGUGGCAUG	30600	13515	AUGCCACUCAAUUUCACGAAGUU	1	1
1781441	AGAUUCGUCUCGCGAAGUAGCGCA	92	47	GUAAUUUUUCACGACGAAUCUAAU	1	2
6650602	GUUGGUUUUGAAUGGCGGACGGUC	27	11	ACAGUCCGCCCGGGGCUCUAACGG	2	1
2472577	AGUUGUUAUCCUAUGGUUGUUCUG	29	4	GUGGAGCCGUGAUGGAUGAAGAGC	5	3
3495498	AUUUCAAGAAGUUGGAGUGGCAUG	34	25	UUCAUGCCACCACAACUCCGUGAA	3	5
7724135	**UGGGCUGUAAAUCGCGAGACGA**	3086	3496	UUUCGCGAUUUACAACCCAUC	9	3
850747	AAUUUCACGGAGUUGUGGUGGCA	5427	1673	AUGCCACUACAAUUUCUUGAAGUU	1	1
6651927	**GUUGUAGAGAUGGCUGCUUGAACA**	5	24			
1509072	AGAACACGAUGAACACAGCAGGUU	19	76	GUCCACUGUGUGUUGUAUUGCUUG	27	6
5564248	**CGUAGAGGCAGCAGCUGCAUA**	32	63	UGCAGUUGUUGUGUCAAGCUU	1	0
8018588	**UUCCAAAUUGUAGGUCGUUUU**	36	7			
4001019	**CCGUUCGCGACGUUCCUGGAG**	52	36	UCGACGAGCGCAGCGUCCGGU	39	9
185087	**AAAUAUAUGACGUUUAGGACA**	128	450	UCCUAAGCGACAUAUAUUUAA	1	2

Figure 10. Predicted fold-back structures for the novel miRNAs that are conserved in other monocots

miRNA s5564248 in switchgrass

UUAA---------- U GA ***A G*** UCC ---------- U
UGCAAGGGG GGU CAA***GCU GA GCAGCAGCUGCAUA*** UGCAAGAAAAU UGGUU U
GCGUUUCCC CCG GUUCGA CU UGUUGUUGACGUGU ACGUUCUUUUA ACCAG C
AACUAUCAAAA U UC A G UGA GUAUCCAU A

Zea mays EST FL355383

```
     C                  A                          C                    A-           UU    U
GG GGCGGCAAGCU GAGGCAGCAACUGCAUA CUUGCAAGAAAA  AUCGGU    UG  U
CC CCGUCGUUCGA CUCUGUUGUUGACGUGU GAACGUUCUUUU  UAGCCA    GC  G
  U                  A                          U                    AG           U-   A
```

Sorghum bicolor CN128779

```
-------------------------          C    GA          A                          U                    ----------          U
                    GCAAGGGG GGU   CAAGCU GAGGCAGCAACUGCAUA  CUUGCAAGAAAAU              UGGUU  U
                    CGUUUCCC CCG   GUUCGA CUCUGUUGUUGACGUGU  GAACGUUCUUUUA              ACCAA  U
AAGUAUCAAAAC              U     UC          A                          U                    CUAGCCAU         G
```

Saccharum hybrid CA107795

```
           U      GA          A                          U                    GUUU      AAC
GCAAGGGG GGU   CAAGCU GAGGCAGCAACUGCAUA  CUUGCAAGAAAAUUG        UCG       C
CGUUUCCC  CCG   GUUCGA CUCUGUUGUUGACGUGU GAACGUUCUUUUAAU        AGC       A
           U      UC          A                          U                    ----         CAU
```

miRNA s8018588 in switchgrass

```
A----              A                                                  C          CC      A   U   C       CAC                U
    UACUCCCU  CGUUCCAAAUUGUAGGUCGUUUUGGUAAAU UAGAUA   UAG UU  UG UAUG      CUAGAUAUA  \
    AUGAGGGA GUAAGGUUUAACAUUCAGUAAAACCGUUUA  AUCUAU   AUU AA  AC  AUAC      GAUCUGUAU  A
GUG              -                                                  A          AU     A   C   A       AUA                U
```

Cenchrus ciliaris EST EB664262

```
.-CUC|             A                  C          UG                 U          AUAUUAU       C  UC       C     A
      CCUCCGUUCC   AAUUGUAGGU  GUUUUG   UUUUCUAGGU  CAUAG            UAUG  A   UAGA AUA  \
      GGAGGCAAGG  UUAAUAUCCA  CAAAAC   AAAAGAUCCA  GUAUU            AUAC  U   AUCU UAU  A
\ ------^              A                  A          CG                 C          GCAAACC       A  GA       A     C
```

miRNA s4001019 in switchgrass

```
CAACC     ACU              ----------      UU     GA        C-          ACUC     C
       CCA      GCCGCGUC         CCG    CGC    CGUUC     UGGAGC       CCG  G
       GGU      CGGCGCAG         GGC    GCG    GCGAG     GCUUCG       GGU  U
CGA---       CGU              UACUU      CU     AC        CA          GCCC     G
```

Zea mays FK969948

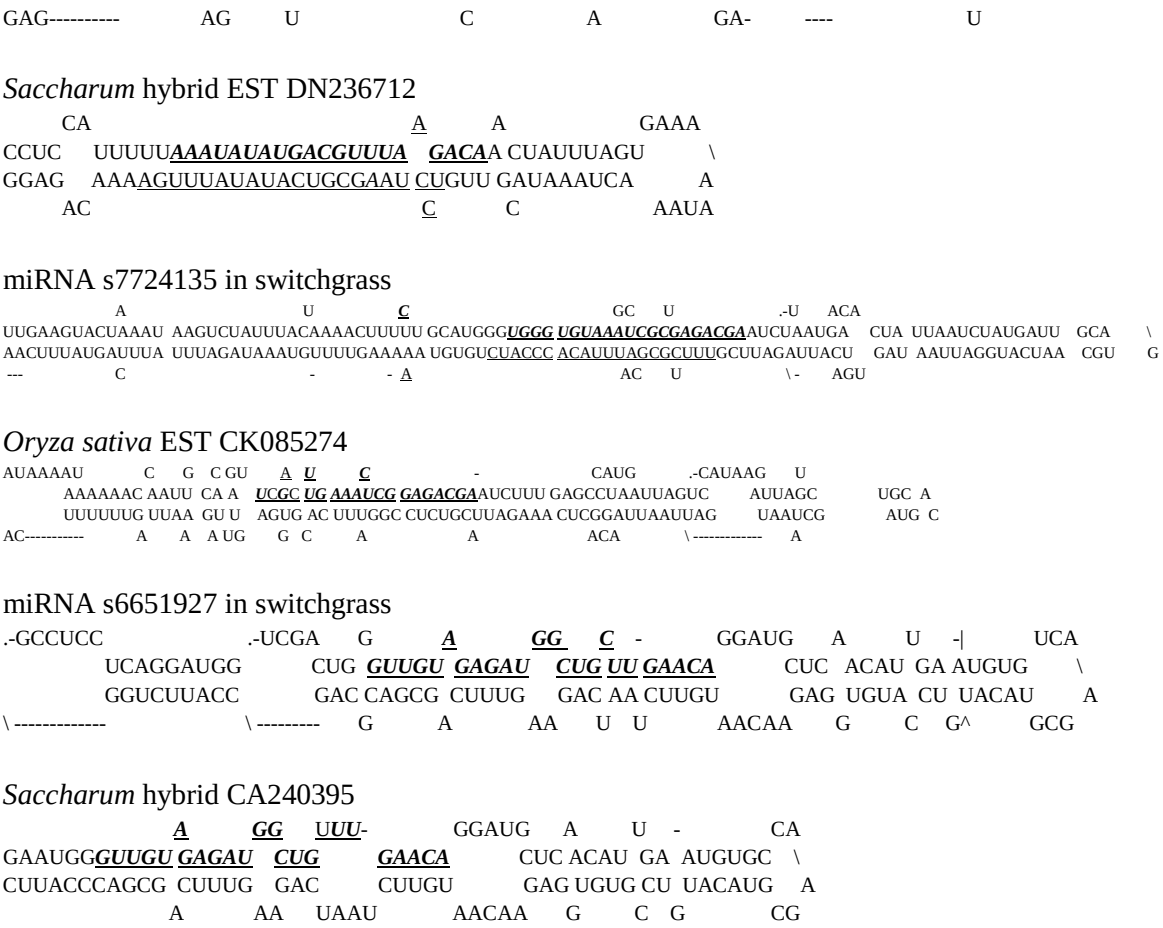
```
          A        .-C          ---------      UU     GA        C-          A     -       C
AGCCCC  ACUGC   GCGUC         CCG    CGC    CGUUC     UGGAGC  CU  CCCG  G
UCGGGG UGGCG   CGCAG         GGC    GCG    GCGAG     GCUUCG  GG GGGU U
          C        \ -          UACUU      CU     AC        CA          -     C       G
```

miRNA s185087 in switchgrass

```
.-CAUGUUCUUUCUC                                                                                  AU      C
                    GUACUCCCUCCAUUCUCAAAUAUAUGACGUUUAGGACAA    GAA  U
                    CAUGAGGGAGGUAAGAGUUUAUAUACUGCAAAUCCUGUU    UUU  A
\ ----------------------------                                                                   GU      A
```

Cenchrus ciliaris EST EB665026

```
AUUGUAUA          GA       C                  A          A          ACA       AAAU           U
             AAAUA    UACU  CCUCCGUUCUU  AAUAUAUG CGUUUAGG     AGU       UAGUUCA A
             UUUGU    AUGA GGAGGCAGGAA UUAUAUAC GCAAAUUC     UCG       AUCAAGU U
```



4.4: Temporal expression analysis of conserved miRNAs

Once a miRNA has been identified, the next step is to determine its temporal and/or spatial expression, which may provide insight into its physiological functions. Many miRNAs and siRNAs are expressed only in certain tissues and cell types, or only during certain developmental stages (Reinhart, *et al.* 2002; Park, *et al.* 2002; Sunkar and Zhu, 2004; Sunkar, *et al*. 2005, 2006). To gain an insight into where these miRNAs are functioning, the expression patterns of 15 conserved miRNAs and 3 other miRNAs (miR528, miR529 and miR5353) that are found only in rice so far, have been analyzed in different tissues and developmental stages. On the basis of signal intensity the abundance of nine miRNAs (miR156, miR160, miR172, miR171, miR167, miR166, miR164,

miR159 and miR319) was relatively higher than that of the other seven miRNAs tested. Several miRNA families, such as miR166, miR159, miR171, miR167, miR160, miR164 and miR398, showed only minor differences in expression levels between tissues (Figure 11). By contrast, some miRNAs showed tissue-specific expression patterns. For instance, the level of miR156 was abundant in both the upper and lower sets of leaves from seedlings but was almost undetectable in similar sets of leaves from adult plant (Figure 11). In contrast, the level of miR172 was abundant in the lower and upper sets of leaves from the adult plant but was almost absent in upper leaves from the seedlings while it was of extremely low abundance in lower leaves of seedlings (Figure 11).

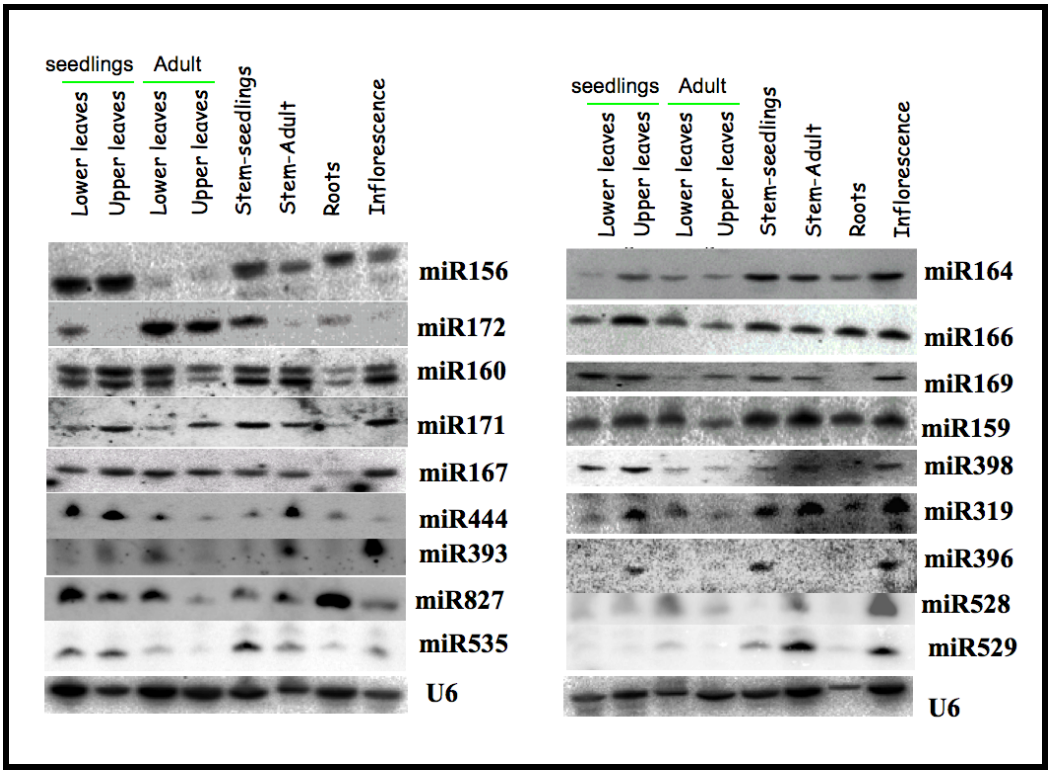


Figure 11: Spatial and temporal expression patterns for eighteen different conserved miRNA families in switchgrass. Modified from previously published paper: Matts, *et al.* (2010) *Journal of Plant Physiology*

miR160 has signals at two different sizes (21 and 22/23-nt) and both are ubiquitously expressed, though not at the same levels in several of the tissues. The signal intensity of the 21-nt form was slightly higher in roots in comparison to the 22/23-nt form, while the signal intensity of 22/23-nt form was slightly higher in the upper leaves of the adult plant. A somewhat equal expression level was observed for the rest of the tissues. . Two different sizes for some miRNAs have been noticed earlier in *Medicago truncatula*, where the 21-nt size has a higher intensity relative to the 22/23-nt size form (Jagadeeswaran, et al. 2009). Two different sizes of the same miRNA family might be derived from the same precursor or from two different precursors in which one is processed as expected (21-nt) and the other precursor is processed to release slightly longer form.

The level of miR393 was abundant only in inflorescences, although it could be detected in stems of adult plants and the upper leaves of both seedlings and adult plants but was almost undetectable in roots, stems and upper leaves of adult plants (Figure 11). MicroRNA, miR319 was abundantly expressed in inflorescence and stems of both seedlings and adult plants and in the upper leaves of seedlings (Figure 11). The expression of miR408, miR528 and miR529 was approximately similar as all these three miRNAs were detected in inflorescence, stem and lower leaves of adult plants (Figure 11). The expression of miR396 was detected only in upper leaves and stems of seedlings and in inflorescence, but was almost absent in other tissues. Although miR171 expression could be detected in all tissues analyzed, it was abundant in upper leaves of both seedlings and adult plants and in stems of seedlings, as well as in inflorescence (Figure 11).

miR444 is a monocot specific miRNA that targets four MADS-box transcription factors in rice (Sunkar, et al. 2005). Interestingly, the expression of miR444 was distinct between leaves of seedlings and adult plants (Figure 11). It was abundant in leaves from seedlings but was low in similar sets of leaves from adult plants. Stems from adult plants showed much higher levels of miR444 than did stems of seedlings. The expression of miR444 seemed to be low in roots and extremely low in inflorescence. miR444 expression in rice was different in comparison to switchgrass. In rice, miR444 had similar expression in the different tissues (Sunkar, et al. 2005). These results suggested a dynamic regulation of several miRNAs in different tissues of switchgrass seedlings and adult plants.

4.5: Analysis of miR395 and miR399 response to nutrient-deprived conditions in switchgrass

Nutrient levels in the soil can vary, and these variations can cause nutrient stress on the plants, depending on how high or low the levels are relative to the normal levels. Plants respond to nutrient deprivation by increasing root growth to access a larger soil volume. Nutrient deprivation can also lead to the slowing of both the growth and development of the plants, and in extreme cases it can even lead to the plant death. It is known that low-levels of phosphate, sulfate, and copper can induce the expression of specific miRNAs in plants. For instance, miR399 is induced under phosphate-deprived conditions whereas miR395 is induced under sulfate-deprived conditions (Pant, *et al*. 2008; Doerner, 2008; Kawashima, *et al*. 2008; Jones-Rhoades, *et al*. 2004). Similarly, miR397, miR398, and miR408 have also been shown to be induced under copper-deprived conditions in *Arabidopsis* and *Medicago truncatula* (Abdel-Ghany and Pilon, 2008; Jagadeeswaran, *et al*. 2009). These miRNAs are widely conserved across different

plant species and anticipated to show similar responses in switchgrass seedlings exposed to sulfate- or phosphate-deprived conditions. However, analysis of miR395 expression in switchgrass indicated its' basal expression is relatively high in seedlings grown on optimal sulfate levels and the miR395 level was only slightly upregulated in sulfate-deprived plants (Figure 12). Similarly, miR399 expression was detected in plants grown with optimal levels of nutrients and only slightly changed under phosphate-deprived conditions (Figure 12).

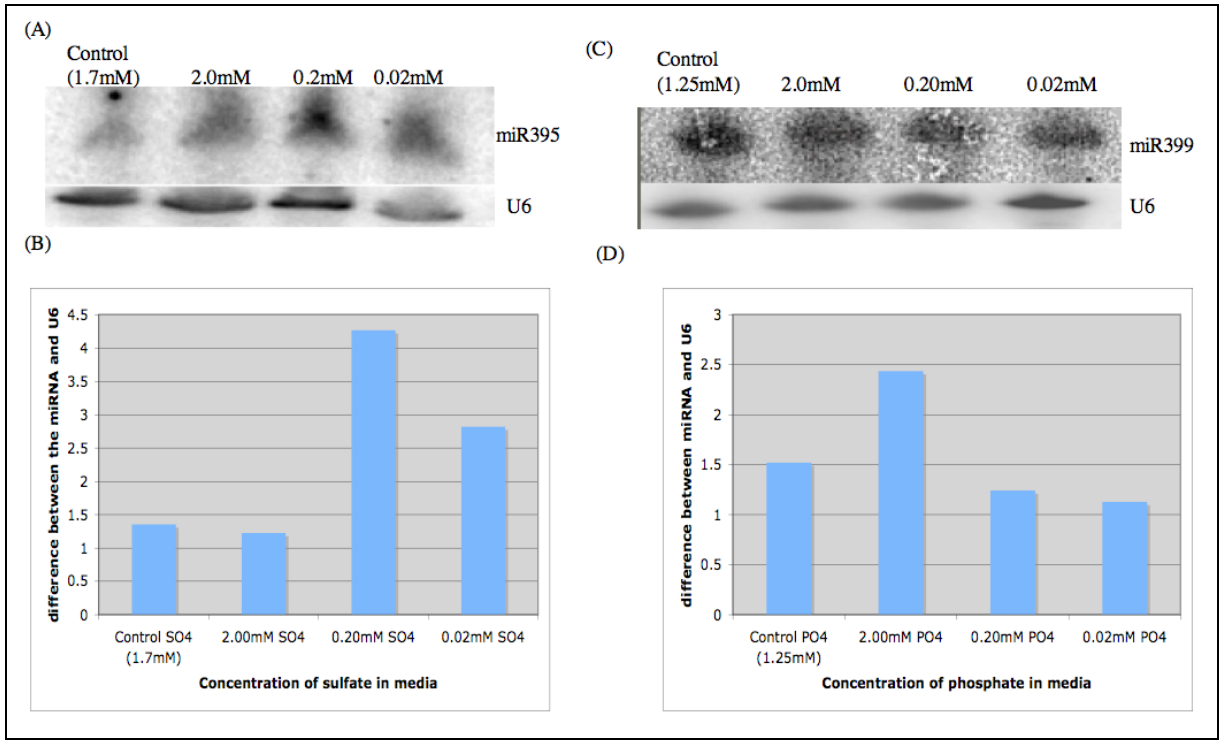


Figure 12. Expression analysis of miR395 (A) and miR399 (C) under sulfate- or phosphate-deprived conditions, respectively. Quantification of altered miR395 (B) and miR399 (D) under stress.

4.6: Temporal expression analysis of novel miRNAs in switchgrass

Five of the novel miRNAs (3496977, 8008250, 7724135, 5564248 and 185087) were analyzed for their temporal expression pattern in switchgrass, using the small RNA blot analysis. Two of the novel miRNAs that were analyzed for temporal expression are switchgrass-specific (s3496977 and s8008250), whereas three others are monocot-specific miRNAs (s7724135, s185087, and s5564248). Many of these novel miRNAs had unique and distinct expression patterns. S7724135 showed much stronger expression relative to the other four novel miRNAs tested. It showed abundant expression in roots, adult stems, upper leaves of both adult plants and seedlings. S3496977 showed low level of expression in all adult tissues analyzed but none in the tissues from seedlings. 8008250 could be detected in upper leaves of adult plants and inflorescence but not in other tissues. Very faint signals were detected for 815087 and 5564248. Thus, like conserved miRNAs, some of the novel miRNAs showed ubiquitous expression whereas some others showed distinct tissue specific expression (Figure 13).

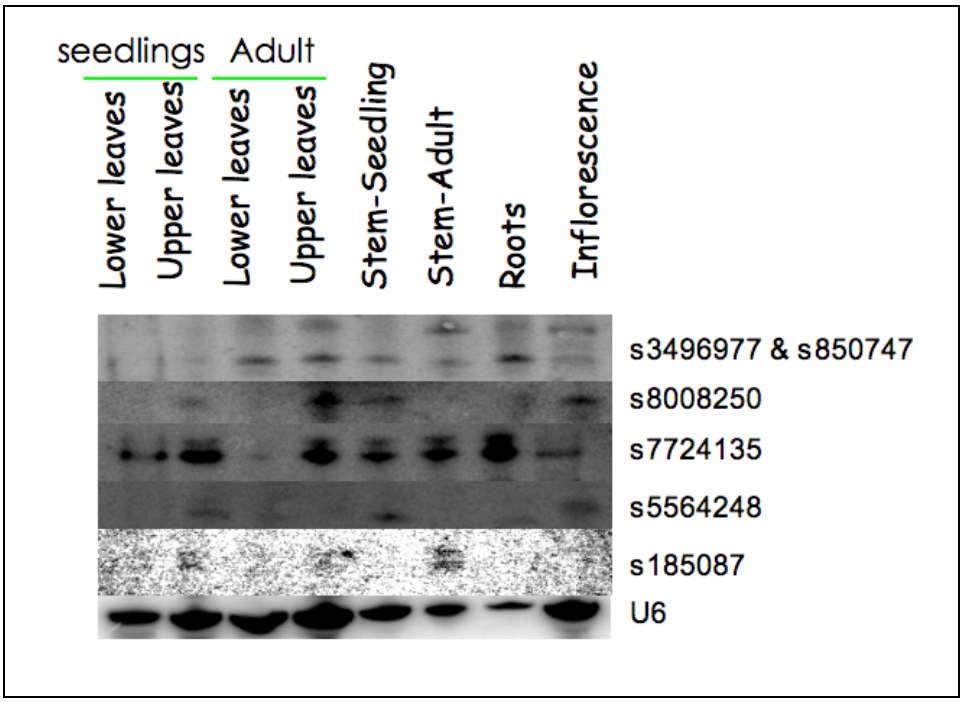


Figure 13: Small RNA blot analysis of novel miRNAs in different switchgrass tissues.

4.7: *Target predictions for conserved miRNAs*

Most plant miRNA sequences display near perfect complementarity with their target mRNAs, and this property has been used to predict potential targets for miRNAs using a computational approach (Rhoades, *et al*. 2002; Sunkar and Zhu, 2004; Jones-Rhoades and Bartel, 2004; Bonnet, *et al.* 2004). To predict potential targets for conserved miRNAs identified in switchgrass, the EST database (NCBI) was searched for switchgrass mRNAs that possess miRNA complementary sites. The alignments of miRNA and their target mRNAs are shown in appendix 6. The 37 predicted targets include homologs of known targets for conserved miRNAs and novel targets. Twelve of the conserved miRNA families are predicted to target transcription factors in *Arabidopsis* (Jones-Rhoades et al., 2006). Similarly in switchgrass, several transcription factor families, including squamosa promoter binding (SBP) transcription factors, MYB

transcription factors, TCP factors, NAC domain-containing transcription factor, auxin response factors (ARFs), Scarecrow-like transcription factors, Apetala-2 (AP2)-like transcription factor, MADS box proteins, and CCAAT-binding factors, were predicted as targets for miR156, miR159, miR319, miR164, miR160/167, miR171, miR172, miR444, and miR169 families, respectively. Other predicted targets include proteins such as transport inhibitor response 1 (an F-box protein) for miR393, laccase for miR397, F-box protein for miR394, DCL-1 for miR162, sulfate transporter for miR395, Argonaute 1-like for miR168, plantacyanin for miR408, ubiquitin-conjugating enzyme for miR399, and transcripts that code for unknown proteins (Table 9).

Table 9. Identified potential targets for conserved miRNA families in switchgrass.

miRNA family	**EST accession number of the predicted target gene**	**Target gene family**
156/157	FE626923; DN143702	SPB-like protein
159	FE65043; GD051711	MYB transcription factor & hypothetical protein
160	FL913173; FL738979; FE606478	Auxin response factor; hypothetical protein; & START domain containing protein
162	FL812781	DCL1
164	FE698722; FL846228	NAC transcription factors
165/166	FE606478; GD002178; FL954559	HD-zip like; HEAT repeat containing protein; & unknown protein
167	DN141844; GD032712; GD007307	Auxin response factor & unknown protein
168	FL904157; FL818096	Argonaute 1 like protein
169	FL965734	CBF (CCAAT-binding factor)-HAP2 like
170/171	FL923024; FL910918	Scarecrow transcription factors
172	FL945982; FL940492; FE642476	AP2 domain containing protein
319	FE603736; FL985594	TCP transcription factor & unknown protein
390	FL692881	Leucine-rich repeat domain containing protein kinase
393	DN143813	F-box protein (TIR1 homolog)
394	FL978450	F-box containing protein
395	FL710917; FL910325	Sulfate transporter & sulfate synthase
397	FL753322	Laccase
399	FL811879; FL997840	Ubiquitin conjugating enzyme protein & transposon
408	FL942386	Plantacyanin
444	FL979804	MADS-box containing transcription factor
528	G052089	Unknown protein

4.8: Validation of selected conserved miRNA targets

To validate the predicted targets, we used 5'-RACE to map the miRNA-guided cleavage site on target mRNAs. Four predicted targets have been validated as genuine targets for miRNAs in switchgrass, using the modified 5'-RACE assay. The validated targets included an NAC transcription factor for miR164, PHB/REV transcription factor for miR166, an SPL transcription factor for miR156, and an AP2-like transcription factor for miR172. All of the sequences showed cleavage between the 10th and 11th nucleotides as expected (Figure 14).

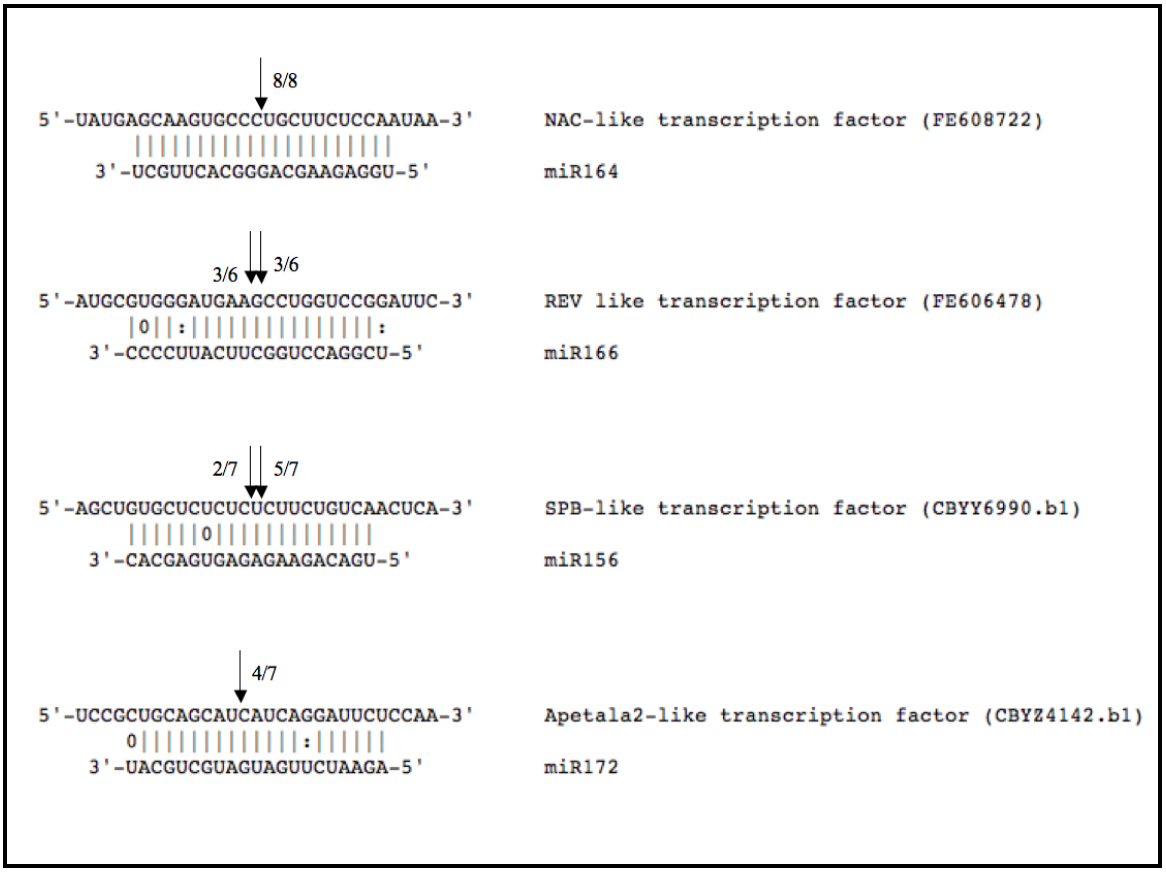


Figure 14: 5'-RACE validation for four conserved miRNA targets in switchgrass. Modified from previously published paper: Matts, *et al*. (2010) *Journal of Plant Physiology*

4.9: Novel miRNA target predictions

Targets for the novel miRNAs were predicted using the same criteria as those for the conserved miRNAs. We were able to predict 28 targets for 12 novel miRNAs in switchgrass (Table 10). While several of these predicted targets are proteins with unknown function or hypothetical proteins, the rest of the target genes encode products that are extremely diverse ranging from glycosyl transferase, sulfotransferase domain containing protein, amino acid permease/amino acid transporter, protease inhibitor, pantothenate kinase, peptide transporter and leucine rich repeat protein. The alignments of all the targets and the novel miRNAs are shown in appendix three.

Table 10: Predicted targets for novel and candidate miRNAs in switchgrass.

Novel miRNA family	**Target EST accession Number**	**Target gene family**
s8008250	FL788822 FL695873 GD04317	Hypothetical protein UDP-glycoronsyl/UDP-glycosyl transferase MTA/SAH nucelosidase
s6815382	GD016940	Sulfotransferase domain containing protein
s3496977 & s850747	FL993113 FL987523, FL764417, & FL846993 FL967210 FL821059 FL754750	Amino acid permease/amino acid transporter Hypothetical proteins Expressed protein Protease inhibitor, seed storage, LTP protein Splicing factor U2AF protein
s1781441	FL854510	Hypothetical protein
s2472577	GD036874 FL787755	Pantothenate kinase Expressed protein
s3495498	FL891811, FL821060, FL952548, & FL787755	Hypothetical proteins
s7724135	FL868007 FL801207 GD035751	Peptide transporter PTR2 Leucine rich repeat protein Unknown protein
s1509072	GD019762 & FL963020	Hypothetical proteins
s5564248	FE657239	Glyscosyl transferase family 1 like protein
s8018588	FL711513	WD40 repeat containing protein
s185087	GD022328 FL978160 FL871007	Expressed protein Acyl-CoA dehydrogenase, mitochondrial precursor TUDOR protein with multiple SNc domains

CHAPTER V

DISUCSSION

Better understanding of the molecular biology of switchgrass will enable us to manipulate the traits for increased biomass production for cellulosic biofuel production.

5.1: Identification of conserved miRNAs in switchgrass

There are a large number of conserved miRNA families across a diverse spectrum of plant species (Axtell and Bartel, 2005; Zhang, et al. 2006; Willmann and Poething, 2007) ranging from monocots to dicots, to ferns and mosses. The conservation of these miRNA families in such diverse plant species indicates conserved functions critical for various developmental pathways and processes such as phase transitions (seedling to adult, and adult to reproductive stages), leaf morphogenesis, meristem boundary formation, leaf development and polarity, lateral root formation and development, flower organ identity and development, reproduction (Jones-Rhoades et al., 2006; Mallory and Vaucheret, 2006) as well as plant tolerance to biotic and abiotic stresses (Shukla et al., 2008; Sunkar, 2010). Three independent small RNA libraries were constructed from three different tissue sources (seedlings, inflorescence and emerging tillers) and subjected to deep sequencing. The analysis of three small RNA libraries revealed the identification of 36 conserved miRNA families, 1 tasiRNA family and 15 novel miRNA families in switchgrass. Besides, the miRNA profiling provided a snapshot of the differences in the miRNA populations between different tissue sources.

Identification of a complete set of conserved miRNA homologs in switchgrass suggests the possibility that the functions described to various miRNAs in other plant species will be applicable for switchgrass also, although there might be slight variations specifically in regulating a number of targets, differential expression and regulation of miRNAs etc.

Within monocots, small RNA populations have been extensively investigated in rice using deep sequencing technology (Sunkar et al., 2008; Lu et al., 2008; Heisel, et al. 2008; Zhu, et al. 2008; Xue, et al. 2009; Fuijoka, et al. 2008; Johnson, et al. 2009). Rice is the most important crop in the world as measured by the portion of calories provided to the human diet, and has served as a model system for monocots, especially for cereals. With the need for biofuel production from plants that are not part of the human diet, there has been a necessity to explore other grass species such as switchgrass, Sorghum, *Brachypodium discantyon* and Miscanthus. In fact, *Brachypodium discantyon* and Sorghum genomes have recently been sequenced (The International Brachypodium Initiative, 2010; Patterson, et al. 2009). Recently attempts were made to analyze the small RNA population of *Brachypodium* (Zhang, et al. 2009; Unver and Budak, 2009; Wei, et al. 2009) and the small RNAs in sorghum have been only computationally predicted so far. Thus, within the biofuel plant species, small RNA-guided gene regulations have not been studied extensively. Our sequencing a total of 20 million small RNA reads is one of the most intensive studies in analyzing small RNA populations in a biofuel plant species.

Within the deep-sequenced small RNA libraries (inflorescence and emerging tillers), miR165/miR166 family had the highest number of reads. The high representation of the miR165/166 family in the inflorescence of switchgrass differed when compared to

Brachypodium. In the libraries for *Brachypodium* constructed by Wei et al., (2009), revealed the highest reads for miR397. The miR165/166 family targets HD-ZIP transcription factors that are essential for specification of adaxial/abaxial (dorso-ventral) polarity. This polarity is established through the polarized expression of *HD-ZIPIII* transcription factors that specify adaxial/upper cell fate (Emery et al. 2003; Juarez et al. 2004; Nogueira et al., 2007). In Arabidopsis and maize, the adaxial-specific expression of *HD-ZIPIII* family members is delineated by the expression pattern of miR166 (Juarez et al. 2004; Kidner and Martienssen 2004; Nogueira et al., 2007). HD-ZIP factors are also implicated in regulating floral development (Jung and Park, 2007). Given the fact that *HD-ZIPIII* family members play diverse roles in leaf and flower development, it is not surprising that their abundance is the highest in inflorescence library. However, the highest abundance in emerging tillers is interesting and this suggests the possibility that miR165/166 -HD-ZIP circuitry is likely to play important roles in tillering.

Within the switchgrass small RNA libraries two family members (TAS3a and TAS3b) of a tasiRNA family were identified. This tasiRNA-production is dependent on the action of miR390 (Allen, et al. 2005). Consistent with this, the abundance of both miR390 and TAS3a in each library showed positive correlations although inflorescences had a higher abundance of both in comparison to the emerging tillers. The tasiRNAs have been shown to play role in leaf development in Arabidopsis (Adenot et al., 2006), but their greater abundance in inflorescence and accumulation in emerging tillers to detectable levels suggests additional roles for the tasiRNAs in plants.

Deep sequencing also recovered two other conserved miRNA families, miR395 and miR399, in switchgrass. This was unique in that these libraries were made from

plants that were grown on optimal nutrient conditions, and it has been shown that miR395 and miR399 are generally detected in plants subjected to nutrient deprived conditions, specifically under sulfate- and phosphate-depleted conditions (Jones-Rhoades and Bartel, 2004; Fujii, et al. 2005; Sunkar et al., 2007; Sunkar, 2010).

Other conserved miRNA families that were identified in the libraries included monocot-specific families such as miR444, miR528, miR529, miR535, miR1318, and miR1432 that have been identified mostly in rice so far. miR444 and miR528 have also been found in *Brachypodium* libraries (Wei, et al. 2009). Of the monocot specific miRNA families, miR528 had the highest abundance in the emerging tillers in comparison to the other families suggesting the possibility that miR528 family could play an important role in tiller emergence. The targets for miR528 could not be predicted in switchgrass, which could be due to the lack of genome or sufficient number of ESTs, but identification of mRNA targets for miR528 could help us understand how miR528 or their targets are playing a role in tiller emergence. Once the miR528 targets are confirmed, strategies to manipulate miR528 or its target gene(s) can be designed for increasing the number of tillers in switchgrass.

5.2: Identification of novel miRNA families in switchgrass

Deep sequencing of small RNA populations led to the identification of 15 novel miRNA families in switchgrass based on cloning the miRNA* sequence and conservation. Interestingly, six of these novel miRNAs are conserved in related monocots, thus can be annotated as monocot-specific miRNAs (Table 8). While several *Brachypodium*-specific miRNAs have been identified (Wei, et al. 2009), but their homologs could not be found in switchgrass small RNA libraries. Similarly, the

homologs for novel miRNAs identified in switchgrass could not be found in Brachypodium, suggesting there are also major differences exists between these two biofuel plant species with respect to the expression of novel miRNAs.

5.3: Temporal expression analyses of conserved miRNAs in switchgrass

Analysis of miRNA expression in *Arabidopsis*, rice and *M. truncatula* revealed many miRNAs expressed in only certain tissues and cell types, only during certain developmental stages, or with altered expression in response to stress (Jagadeeswaran et al., 2009b; Lu et al., 2005; Sunkar and Zhu, 2004; Sunkar et al., 2005; 2006). Furthermore, previous reports showed the conserved miRNAs with divergent expression patterns in different plant species (Jagadeeswaran et al., 2009b; Lu et al., 2005; Subramanian et al., 2008).

Eighteen conserved miRNA families were examined for their expression patterns in different tissues of switchgrass (Figure 11). Most of the miRNA families were ubiquitously expressed in the different tissues of switchgrass, whereas some miRNAs showed very distinct tissue-specific expression patterns. Several of the miRNA families, such as miR159, miR160, miR164, miR166, miR167, miR171, and miR398 showed minor differences in their expression levels between the different tissues.

For instance, miR156 is abundantly expressed in the lower and upper leaves of the seedlings, but was almost undetectable in similar sets of leaves in the adult plant. In contrast, miR172 was abundantly expressed in the lower and upper leaves of the adult plant, but was almost undetectable in the upper leaves of the seedlings, and had diminished levels in the lower leaves of the seedlings. This opposite pattern of expression for miR156 and miR172 is consistent with their reported roles in *Arabidopsis*

(Wang, et al. 2009). Transcripts encoding Squamosa promoter binding factor like proteins (SPLs) and AP2-like factors are targeted by miR156 and miR172 respectively, and these genes have been shown to play an important role in phase changes in *Arabidopsis*. *Arabidopsis* miR156 plays an indispensable role in controlling the phase transitions form juvenile phase-to-adult phase by targeting SPL transcription factors (Wang, et al. 2009; Wu, et al. 2009; Yamaguchi, et al. 2009). In contrast, miR172 appears to play a role in the control of vegetative-phase to reproductive-phase transitions in *Arabidopsis*. Surprisingly, in switchgrass, miR172 was almost undetectable in the flower tissue, whereas miR172 was abundantly expressed in the flowers of *Arabidopsis* (Chen, 2004).

miR160 and miR167 could be detected in all tissues examined with the exception of root tissue in which the expression level is below the detection limit. Both miR160 and miR167 are known to target auxin response factor (ARF) transcription factor family members. These transcription factors are known to play key roles in seed germination (Liu, et al. 2007), leaf development (Mallory, et al. 2005), and floral development (Wu, et al. 2006).

miR319 is abundantly expressed in the flowers, and in the upper leaves and stems of both the seedlings and mature plants. It is known that miR319 plays an important role in leaf morphogenesis by targeting TCP factors in *Arabidopsis* (Palatnik, et al. 2003) and tomato (Ori, et al. 2007). However, its detection in the flowers and stems strongly suggest additional roles for miR319 in switchgrass. One of the roles could be flower development, as it has been recently shown that miR319 targets TCP factors that are critical for proper floral development (anthers and stamens) in *Arabidopsis* (Nag, et al.

2009). miR159 was abundantly expressed in inflorescence, roots, and stems of both the seedlings and adults. This is consistent with a recent study that reported abundant expression of miR159 in several tissues of *Medicago truncatula* (Jagadeeswaran, et al. 2009).

miR171 expression could be detected in all the tissues analyzed, although most abundant expression was evident in the upper leaves of both the seedlings and mature plants, the stems of the seedlings, as well as in the flowers. This differs from rice in which miR171 showed abundant expression in all the tissues analyzed (Sunkar, et al. 2005). The expression levels of miR171 in switchgrass is rather high when compared to *Medicago,* in which miR170/miR171 showed only low level expression in all the tissues analyzed (Jagadeeswaran, et al. 2009).

Interestingly, miR444 showed distinct expression patterns in the leaves of both the seedlings and adult plants. It had low expression in the roots, while expression was even lower in the inflorescence in switchgrass. This expression pattern differs with rice in which miR444 showed similar expression levels in different tissues (Sunkar, et al. 2005). These results suggest a very dynamic regulation of miR444 expression in different tissues in switchgrass. miR444 is a monocot specific miRNA that targets four MADS-box transcription factors in rice (Sunkar, et al. 2005; Li et al., 2010). Also it was shown that miR444 family regulates at multiple sites on these MADS box genes in rice (Li et al., 2010). *MADS*-box genes are known to play a critical role in determining the organ specificity during flower development in Arabidopsis whereas in rice these genes suggested to play important roles in meristem identity, formation of the dehiscence zone, fruit ripening, embryo development as well as development of vegetative organs

(Kaufmann et al., 2009; Arora et al., 2007; Li et al., 2010). Taken together, on the one hand there are similarities in expression patterns of similar miRNAs in different tissues of switchgrass and rice but there are also differences for some of the miRNAs. Target gene expression analysis could provide better insights in such cases.

5.4: The response of miR395 and miR399 to different concentrations of sulfate and phosphate

It has been shown that miR395 and miR399 are up regulated in response to low sulfate and low phosphate conditions, respectively in other plant species (Jones-Rhoades and Bartel, 2004; Fujii, *et al*. 2005; and Jagadeeswaran, *et al*. 2009). However, miR395 and miR399 in switchgrass are constitutively expressed at relatively higher levels (also reflected by their frequency in inflorescence and emerging tillers libraries) and the up-regulation was not significant in seedlings grown on low-sulfate when compared to control plants (figure 12), and low-phosphate when compared to control plants (figure 12). Thus, the regulation of miR395 and miR399 in switchgrass is different in comparison to *Arabidopsis thaliana, Medicago truncatula,* and rice. One plausible explanation for non-regulation of these miRNAs in switchgrass is that these plants have been grown on wastelands and marginal soils and perhaps over the time these plants have developed an adaptive mechanism by constitutively turning on such key regulators that can confer tolerance to the plant species.

5.5: Temporal expression analyses of novel miRNAs

Five novel miRNAs that were analyzed showed unique and distinct expression patterns. It was known that occasionally, sequence-based expression profiling and small RNA blot analysis does not correlate (Reddy et al., 2009). Some of the switchgrass novel miRNA analyses revealed no correlations between these two assays. For example,

miRNA s3496977 had the highest number of reads in the inflorescence library (2,600 normalized reads), whereas s7724135 had 265 reads in the inflorescence library, representing a ten-fold difference between these two miRNAs. However, small RNA blot analyses revealed an opposite trend, i.e., s7724135, which only had 265 reads in the inflorescence library is more abundant in the inflorescence whereas s3496977 had a fairly low expression in the inflorescence. The biased cloning could be attributed to the differences in the 5'- and 3'-end nucleotides, which are involved in ligating with the 5' and 3' adapters, respectively, or formation of secondary structures or adoption of a structure that prevents the exposure of 5' or 3' ends of the miRNA (Reddy et al., 2009). Another possible reason for the difference could be the rate in which the mature miRNA is incorporated into the RISC and once the miRNA is incorporated into the RISC complex, it is shielded by the surrounding proteins, and is lost during the RNA extraction.

5.6: Identification of targets for conserved microRNAs in switchgrass.

Targets have been predicted for majority of the conserved miRNAs in switchgrass (Table 9 and Appendix 2). The predicted targets include homologs of known targets for conserved miRNAs, as well as novel targets. Twelve of the conserved miRNA families have been predicted to target transcription factor families in *Arabidopsis* (Jones-Rhoades, et al. 2006). Consistent with this, in switchgrass the transcription factor families such as squamosa promoter binding (SBP) transcription factors, MYB transcription factors, TCP transcription factors, NAC-domain containing transcription factors, auxin response factors (ARFs), CCAAT-binding factors, scarecrow-like transcription factors, Apetala2-like transcription factors, MADS-box proteins have been predicted as targets for miR156,

miR159, miR319, miR164, miR160/167, miR169, miR171, miR172 and miR444 families respectively. miR395 has been predicted to target a sulfate transporter (FL710917), which is a conserved target.

Other predicted targets include proteins such as TIR1 (transport inhibitor response 1, an F-box protein) transcript for miR393, Dicer-like 1 for miR162, Argonaute 1-like for miR168, laccase for miR397, plastocyanin for miR408, and transcripts that code for unknown proteins. In addition to the conserved targets for conserved miRNAs, recent studies indicated that some conserved miRNAs may have non-conserved targets in different plant species (German, *et al*. 2008; Li et al., 2010). Consistent with this suggestion, based on sequence complementarity we have identified several novel targets for conserved miRNAs in switchgrass. A few such examples are trihelix-domain containing transcription factor (FL9210909) for miR159, and a HEAT-repeat containing protein (GD002178) for miR166. mR395 also showed extensive complementarity with a bifunctional 3'-phosphoadenosine 5-phosphosulfate synthetase (FL910325), suggesting that it might be a potential non-conserved target in switchgrass. The bifunctional sulfate synthetase seems to be a conserved target in grasses, as it was also predicted as a target of miR395 in *Brachypodium* (Wei, et al. 2009).

The potential target for miR159 (FL9210909), a trihelix-domain containing transcription factor, is a plant specific class of transcription factors that are known to interact with the promoter regions of light responsive genes in the nucleus (Smalle, et al. 1998). Targets could not be predicted for some of the conserved miRNAs and this is not surprising given the fact that the target predictions in switchgrass only dependent on the available ESTs in the database.

5.7: Validation of predicted conserved miRNA targets using modified 5'-RACE assay

Four predicted targets were validated as genuine miRNA targets in switchgrass, using modified 5'RACE assays. It is known that the RISC complex cleaves the targeted mRNA between the 10^{th} and 11^{th} nucleotides of the recognition sequence. The validated targets include NAC transcription factor for miR164, PHB/REV transcription factor (a homolog for an HD-Zip transcription factor) for miR166, SPL transcription factor for miR156 and AP2-like transcription factor for miR172. Majority of the sequences showed cleavage between the 10^{th} and 11^{th} nucleotides. It has also been recently shown that plant microRNAs can also act in translational repression in addition to mRNA cleavage (Voinnet and Brodersen, 2010) suggesting that analyzing the regulation at the protein level also becomes important in plants.

5.8: Identification of targets for novel miRNAs in switchgrass

Predicting targets for many species-specific miRNAs in both *Arabidopsis* and rice has been not very successful (Fahlgren, *et al*. 2007; Lu, *et al.* 2008; Sunkar, *et al*. 2008). We were able to predict 28 targets for 12 novel miRNAs and targets could not be predicted for 3 novel miRNAs in switchgrass (Table 10). Unlike the conserved miRNA targets most of which are transcription factors, novel miRNAs are targeting a wide spectrum of genes involved diverse physiological processes. For instance, sugar transfer proteins are targets of miRNAs, s8008250 and s5564248. An UDP-glycoronsyl/UDP-glycosyl transferase protein is a target of s8008259; while an glyscosyl transferase family 1 like protein is a target of s5564248. These types of proteins catalyze the transfer of sugars to numerous different substrates (Ross, *et al*. 2001; Bowles, *et al*. 2006). The

determination of the substrate for both of these targets, will determine what pathway or other proteins they interact with.

Another potential target is an aldehyde oxidase family member (miRNA family s7470078). This broad class of enzymes is highly conserved from microorganisms to plants and animals and is shown to function in the oxidation of various aldehydes (Sekimoto, *et al*. 1997). Within plants, aldehyde oxidases play a role in the biosynthesis of the plant hormones abscisic acid and indole-3-acetic acid (Schwartz, *et al*. 2003; Normanly, 2010). It has been shown that these hormones regulate miRNA expression in plants (Yang, *et al*. 2006; Zhao, *et al.* 2009).

Another potential target is pantothenate kinase for family s2472577. This enzyme is essential for the coenzyme A biosynthesis pathway, where it catalyzes the first step of the reaction (Kupke, *et al*. 2003; Tilton, *et al.* 2006). Coenzyme A, is an essential for synthesis and oxidation of fatty acids, and the oxidation of pyruvate in the citric acid cycle, therefore its proper regulation is essential. The remaining potential miRNA targets include splicing factors, transport proteins, sulfate transfer proteins, co-chaperone proteins, and nuceloidases. The broad range of targets for the novel miRNAs in switchgrass suggest that that miRNAs could also play important roles in most cellular biological processes in switchgrass.

The future of switchgrass for use as a biofuel stock species will rely upon unlocking the molecular circuitry controlling biomass production (prolonged vegetative phase, increased foliar number and size, plant height, tiller numbers, etc.) and tolerance to the abiotic (drought, heat, cold, and nutrient stress) and biotic stresses. Despite the

increasing importance of switchgrass as an energy crop, little is known about the basic biology of the traits that make switchgrass a useful biofuel plant species. Recent studied in diverse plant species indicate that most of the traits that make switchgrass as a biofuel are at least partly under the control of miRNA-guided gene regulation (Jones-Rhoades et al., 2006; Mallory and Vaucheret, 2006; Sunkar et al., 2007; Sunkar, 2010). A better understanding of such processes in switchgrass will aid in manipulating the biomass production or stress tolerance in this important biofuel plant species.

Present study provided some hints as to what miRNAs might be logical to use for improving biomass production. Certainly, miR156 is the prime candidate for overexpression studies aiming at improving biomass production. Relatively higher abundance of miR528 is worth considering for future manipulation of increasing the tiller numbers in switchgrass, although identifying its target genes is top priority. Based on the target gene function, research strategies can be devised as to whether overexpression of miR528 or overexpression of miR528-resistant target gene needs to be undertaken. In summary, this work provided a glimpse of miRNAs and their targets in switchgrass. Identification of conserved as well as novel miRNA families and RNA targets for most of miRNAs in switchgrass can serve as a foundation for future further characterization of miRNA regulatory networks as well as functional genomics approaches in switchgrass.

CHAPTER VI

CONCLUSIONS

Identification of conserved and novel miRNAs

Using computational approach, homologs and fold back structures were predicted for thirteen of the highly conserved miRNA families and three of the monocot-specific miRNA families in switchgrass. With the use of deep sequencing, thirty five miRNA families, some of which are highly conserved miR156, miR159, miR160, miR162, miR164, miR165/166, miR167, miR169, miR170/171, miR172, miR319, miR390, miR393, miR394, miR395, miR396, miR397, miR398, miR399, and miR408, and some others that are conserved in specific lineages or closely related plant species (miR437, miR444, miR528, miR529, miR535, miR827, miR845, miR894, miR1318, miR1432, miR2118, miR2275, miR2910, miR2915 and miR2916) have been identified in switchgrass. Most of these conserved miRNAs are also highly expressed as determined by their frequency in the small RNA libraries as well as expression-based analyses (small RNA blot assay). Recent studies clearly established that plant small RNA population consist not only conserved miRNAs but also lineage-specific or species-specific miRNAs. Deep sequencing of small RNA libraries revealed the existence of fifteen novel miRNA families of which six are conserved at least in one another monocot species (monocot-specific miRNAs) and nine are annotated as switchgrass-specific because their homologues could not be found in other species as of now.

Thus, it is conclusively shown that the switchgrass, a model cellulosic biofuel plant species, has a dynamic population of conserved and novel miRNAs.

Temporal expression analyses of conserved and novel miRNAs

The temporal expression patterns were determined for 18 of the conserved miRNAs in eight different tissues of switchgrass. The expression patterns of several conserved miRNA families differ greatly within the tissues in switchgrass and also in comparison to the other plants. The expression analyses of five novel miRNAs, three of which are conserved at least in one another monocot and two other miRNAs that are switchgrass-specific have been analyzed in eight different tissues. Their expression pattern also differed greatly between different tissues. Both miR395 and miR399, in general, are induced in response to sulfate- and phosphate-deprivation in Arabidopsis, rice and *Medicago truncatula*. However, these two miRNAs are constitutively expressed at higher levels in switchgrass. Non-regulation of miR395 and miR399 could have been the result of adaptation of switchgrass to marginal and infertile soils.

Predicted targets for conserved and novel miRNAs and their validation

The targets for a majority of the conserved miRNA families have been identified. These include squamosa promoter binding (SBP) transcription factors, MYB transcription factors, TCP factors, NAC domain-containing transcription factor, auxin response factors (ARFs), Scarecrow-like transcription factors, Apetala-2 (AP2)-like transcription factor, MADS box proteins, and CCAAT-binding factors, were predicted as targets for miR156, miR159, miR319, miR164, miR160/167, miR171, miR172, miR444, and miR169 families, respectively. Other predicted targets include proteins such as transport inhibitor

response 1 (an F-box protein) for miR393, laccase for miR397, F-box protein for miR394, DCL-1 for miR162, sulfate transporter for miR395, Argonaute 1-like for miR168, plantacyanin for miR408, ubiquitin-conjugating enzyme for miR399, and transcripts that code for unknown proteins. A few of the non-conserved targets for conserved miRNAs trihelix-domain containing transcription factor (FL9210909) for miR159, and a HEAT-repeat containing protein (GD002178) for miR166 and a bifuncational sulfate synthetase for miR395. Four of the predicted targets for conserved miRNAs in switchgrass have been validated using modified 5'-RACE assay. The validated targets are NAC transcription factor for miR164, PHB/REV transcription factor (a homolog for an HD-Zip transcription factor) for miR166, SPL transcription factor for miR156 and AP2-like transcription factor for miR172. Additionally, predicted 28 mRNAs as targets for 12 novel miRNAs in switchgrass.

REFERENCES

Addo-Quaye, C., Eshoo, T.W., Bartel, D.P., and Axtell, M.J. (2008). Endogenous siRNA and miRNA targets identified by sequencing of the Arabidopsis degradome. Curr Biol 18, 758-762.

Adenot, X., **Elmayan, T., Lauressergues, D., Boutet, S., Bouche, N., Gasciolli, V., and Vaucheret, H.** (2006). DRB4-dependent TAS3 trans-acting siRNAs control leaf morphology through AGO7. Curr Biol 16, 927-932

Allen, E., Xie, Z., Gustafson, A.M., and Carrington, J.C. (2005). microRNA-directed phasing during trans-acting siRNA biogenesis in plants. Cell 121, 207-221**.**

Ambros, V. (2003). MicroRNA pathways in flies and worms: growth, death, fat, stress, and timing. Cell 113, 673-676.

Ambros, V. (2004). The functions of animal microRNAs. Nature 431, 350-355.

Arazi, T., Talmor-Neiman, M., Stav, R., Riese, M., Huijser, P., and Baulcombe, D.C. (2005). Cloning and characterization of micro-RNAs from moss. Plant J 43, 837-848.

Aukerman, M.J., and Sakai, H. (2003). Regulation of flowering time and floral organ identity by a MicroRNA and its APETALA2-like target genes. Plant Cell 15, 2730-2741.

Axtell, M.J., and Bartel, D.P. (2005). Antiquity of microRNAs and their targets in land plants. Plant Cell 17, 1658-1673.

Barakat, A., Wall, P.K., Diloreto, S., Depamphilis, C.W., and Carlson, J.E. (2007). Conservation and divergence of microRNAs in Populus. BMC Genomics 8, 481.

Bollman, K.M., Aukerman, M.J., Park, M.Y., Hunter, C., Berardini, T.Z., and Poethig, R.S. (2003). HASTY, the Arabidopsis ortholog of exportin 5/MSN5, regulates phase change and morphogenesis. Development 130, 1493-1504.

Borsani, O., Zhu, J., Verslues, P.E., Sunkar, R., and Zhu, J.K. (2005). Endogenous siRNAs derived from a pair of natural cis-antisense transcripts regulate salt tolerance in Arabidopsis. Cell 123, 1279-1291.

Boualem, A., **Laporte, P., Jovanovic, M., Laffont, C., Plet, J., Combier, J.P., Niebel, A., Crespi, M., and Frugier, F.** (2008). MicroRNA166 controls root and nodule development in Medicago truncatula. Plant J 54(5): 876-887

Bouton, J.H. (2007). Molecular breeding of switchgrass for use as a biofuel crop. Curr Opin Genet Dev 17, 553-558.

Brodersen, P., and Voinnet, O. (2009). Revisiting the principles of microRNA target recognition and mode of action. Nat Rev Mol Cell Biol 10, 141-148.

Carra, A., Mica, E., Gambino, G., Pindo, M., Moser, C., Pe, M.E., and Schubert, A. (2009). Cloning and characterization of small non-coding RNAs from grape. Plant J 59, 750-763..

Chan, S.W., Zilberman, D., Xie, Z., Johansen, L.K., Carrington, J.C., and Jacobsen, S.E. (2004). RNA silencing genes control de novo DNA methylation. Science 303, 1336.

Chekulaeva, M., and Filipowicz, W. (2009). Mechanisms of miRNA-mediated post-transcriptional regulation in animal cells. Curr Opin Cell Biol 21, 452-460.

Cho, S.H., Addo-Quaye, C., Coruh, C., Arif, M.A., Ma, Z., Frank, W., and Axtell, M.J. (2008). Physcomitrella patens DCL3 is required for 22-24 nt siRNA acculumation, suppression of retrotransposon-derived transcripts, and normal development. Plant Cell. 19(9): 2705-2718

Dong, Z., Han, M.H., and Fedoroff, N. (2008). The RNA-binding proteins HYL1 and SE promote accurate in vitro processing of pri-miRNA by DCL1. Proc Natl Acad Sci U S A 105, 9970-9975.

Fahlgren, N., Montgomery, T.A., Howell, M.D., Allen, E., Dvorak, S.K., Alexander, A.L., and Carrington, J.C. (2006). Regulation of AUXIN RESPONSE FACTOR3 by TAS3 ta-siRNA affects developmental timing and patterning in Arabidopsis. Curr Biol 16, 939-944.

Fahlgren, N., Howell, M.D., Kasschau, K.D., Chapman, E.J., Sullivan, C.M., Cumbie, J.S., Givan, S.A., Law, T.F., Grant, S.R., Dangl, J.L., and Carrington, J.C. (2007). High-throughput sequencing of Arabidopsis microRNAs: evidence for frequent birth and death of MIRNA genes. PLoS One 2, e219.

Fujii, H., Chiou, T.J., Lin, S.I., Aung, K., and Zhu, J.K. (2005). A miRNA involved in phosphate-starvation response in Arabidopsis. Curr Biol 15, 2038-2043.

Gandikota, M., Birkenbihl, R.P., Hohmann, S., Cardon, G.H., Saedler, H., and Huijser, P. (2007). The miRNA156/157 recognition element in the 3' UTR of the Arabidopsis SBP box gene SPL3 prevents early flowering by translational inhibition in seedlings. Plant J 49, 683-693.

Hamilton, A.J., and Baulcombe, D.C. (1999). A species of small antisense RNA in posttranscriptional gene silencing in plants. Science 286, 950-952.

Hamilton, A., Voinnet, O., Chappell, L., and Baulcombe, D. (2002). Two classes of short interfering RNA in RNA silencing. EMBO J. 21(17): 4671-4679

Han, M.H., Goud, S., Song, L., and Fedoroff, N. (2004). The Arabidopsis double-stranded RNA-binding protein HYL1 plays a role in microRNA-mediated gene regulation. Proc Natl Acad Sci U S A 101, 1093-1098.

Hinas, A., Reimegard, J., Wagner, E.G., Nellen, W., Ambros, V.R., and Soderbom, F. (2007). The small RNA repertoire of Dictyostelium discoideum and its regulation by components of the RNAi pathway. Nucleic Acids Res 35, 6714-6726.

Hsieh, L.C., Lin, S.I., Shih, A.C., Chen, J.W., Lin, W.Y., Tseng, C.Y., Li, W.H., and Chiou, T.J. (2009). Uncovering small RNA-mediated responses to phosphate deficiency in Arabidopsis by deep sequencing. Plant Physiol 151, 2120-2132.

Ibrahim, F., Rymarquis, L.A., Kim, E.J., Becker, J., Balassa, E., Green, P.J., and Cerutti, H. (2010). Uridylation of mature miRNAs and siRNAs by the MUT68 nucleotidyltransferase promotes their degradation in Chlamydomonas. Proc Natl Acad Sci U S A 107, 3906-3911.

Itaya, A., Bundschuh, R., Archual, A.J., Joung, J.G., Fei, Z., Dai, X., Zhao, P.X.,

Tang, Y., Nelson, R.S., and Ding, B. (2008). Small RNAs in tomato fruit and leaf development. Biochim Biophys Acta 1779, 99-107**.**

Jagadeeswaran, G., Saini, A., and Sunkar R. (2009a). Biotic and abiotic stress down-regulate miR398 expression in Arabidopsis. Planta 229, 1009-1014

Jagadeeswaran, G., Zheng, Y., Li, Y.F., Shukla, L.I., Matts, J., Hoyt, P., Macmil, S.L., Wiley, G.B., Roe, B.A., Zhang, W., and Sunkar, R. (2009b). Cloning and characterization of small RNAs from Medicago truncatula reveals four novel legume-specific microRNA families. New Phytol 184, 85-98.

Jeong, D.H., German, M.A., Rymarquis, L.A., Thatcher, S.R., and Green, P.J. (2010). Abiotic stress-associated miRNAs: detection and functional analysis. Methods Mol Biol 592, 203-230

Jones-Rhoades, M.W., and Bartel, D.P. (2004). Computational identification of plant microRNAs and their targets, including a stress-induced miRNA. Mol Cell 14, 787-799.

Jones-Rhoades, M.W., Bartel, D.P., and Bartel, B. (2006). MicroRNAS and their regulatory roles in plants. Annu Rev Plant Biol 57, 19-53.

Khvorova, A., Reynolds, A., and Jayasena, S.D. (2003). Functional siRNAs and miRNAs exhibit strand bias. Cell. 115(2): 505

Kim, S., Yang, J.Y., Xu, J., Jang, I.C., Prigge, M.J., and Chua, N.H. (2008). Two cap-binding proteins CBP20 and CBP80 are involved in processing primary MicroRNAs. Plant Cell Physiol 49, 1634-1644.

Ko, J.H., Prassinos, C., and Han, K.H. (2006). Developmental and seasonal expression of PtaHB1, a Populus gene encoding a class III HD-Zip protein, is closely associated with secondary growth and inversely correlated with the level of microRNA (miR166). New Phytol 169, 469-78

Kurihara, Y., and Watanabe, Y. (2004). Arabidopsis micro-RNA biogenesis through Dicer-like 1 protein functions. Proc Natl Acad Sci U S A 101, 12753-12758.

Kurihara, Y., Takashi, Y., and Watanabe, Y. (2006). The interaction between DCL1 and HYL1 is important for efficient and precise processing of pri-miRNA in plant microRNA biogenesis. Rna 12, 206-212.

Laubinger, S., Sachsenberg, T., Zeller, G., Busch, W., Lohmann, J.U., Ratsch, G., and Weigel, D. (2008). Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. Proc Natl Acad Sci U S A 105, 8795-8800.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-854.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. Embo J 23, 4051-4060.

Lelandais-Briere, C., Sorin, C., Declerck, M., Benslimane, A., Crespi, M., and Hartmann, C. (2010). Small RNA diversity in plants and its impact in development. Curr Genomics. 11(1): 14-23

Li, J., Yang, Z., Yu, B., Liu, J., and Chen, X. (2005). Methylation protects miRNAs and siRNAs from a 3'-end uridylation activity in Arabidopsis. Curr Biol 15, 1501-1507.

Liu, B., Li, P., Li, X., Liu, C., Cao, S., Chu, C., and Cao, X. (2005). Loss of function

of OsDCL1 affects microRNA accumulation and causes developmental defects in rice. Plant Physiol 139, 296-305.

Liu, B., **Chen, Z., Song, X., Liu, C., Cui, X., Zhao, X., Fang, J., Xu, W., Zhang, H., Wang, X., Chu, C., Deng, X., Xue, Y., and Cao, X.** (2007). Oryza sativa dicer-like4 reveals a key role for small interfering RNA silencing in plant development. Plant Cell. 19(9): 2705-2718

Llave, C., Kasschau, K.D., Rector, M.A., and Carrington, J.C. (2002). Endogenous and silencing-associated small RNAs in plants. Plant Cell 14, 1605-1619.

Lu, C., Kulkarni, K., Souret, F.F., MuthuValliappan, R., Tej, S.S., Poethig, R.S., Henderson, I.R., Jacobsen, S.E., Wang, W., Green, P.J., and Meyers, B.C. (2006). MicroRNAs and other small RNAs enriched in the Arabidopsis RNA-dependent RNA polymerase-2 mutant. Genome Res 16, 1276-1288.

Lu, C., Jeong, D.H., Kulkarni, K., Pillay, M., Nobuta, K., German, R., Thatcher, S.R., Maher, C., Zhang, L., Ware, D., Liu, B., Cao, X., Meyers, B.C., and Green, P.J. (2008). Genome-wide analysis for discovery of rice microRNAs reveals natural antisense microRNAs (nat-miRNAs). Proc Natl Acad Sci U S A 105, 4951-4956.

Mallory, A.C., and Vaucheret, H. (2006). Functions of microRNAs and related small RNAs in plants. Nat Genet 38 Suppl, S31-36.

Matts, J., Jagadeeswaran, G., Roe, B.A., and Sunkar, R. (2010). Identification of microRNAs and their targets in switchgrass, a model biofuel plant species. J Plant Phy 167, 896-904

Millar, A.A., and Waterhouse, P.M. (2005). Plant and animal microRNAs: similarities and differences. Funct Integr Genomics 5, 129-135.

Montgomery, T.A., Howell, M.D., Cuperus, J.T., Li, D., Hansen, J.E., Alexander, A.L., Chapman, E.J., Fahlgren, N., Allen, E., and Carrington, J.C. (2008a). Specificity of ARGONAUTE7-miR390 interaction and dual functionality in TAS3 trans-acting siRNA formation. Cell 133, 128-141.

Montgomery, T.A., Yoo, S.J., Fahlgren, N., Gilbert, S.D., Howell, M.D., Sullivan, C.M., Alexander, A., Nguyen, G., Allen, E., Ahn, J.H., and Carrington, J.C. (2008b). AGO1-miR173 complex initiates phased siRNA formation in plants. Proc Natl Acad Sci U S A 105, 20055-20062.

Nag, A., and Jack, T. (2010). Sculpting the flower: the role of microRNAs in flower development. Curr Top Dev Biol. 91: 349-378

Nag, A., King, S., and Jack, T. (2009). miR319a targeting of TCP4 is critical for petal growth and development in Arabidopsis. Proc Natl Acad Sci U S A 106, 22534-22539.

Nakayashiki, H., and Nguyen, Q.B. (2008). RNA interference: roles in fungal biology. Curr Opin Microbiol 11, 494-502.

Nakayashiki, H., Kadotani, N., and Mayama, S. (2006). Evolution and diversification of RNA silencing proteins in fungi. J Mol Evol 63, 127-135.

Nelson, R.G., Ascough, J.C., 2nd, and Langemeier, M.R. (2006). Environmental and economic analysis of switchgrass production for water quality improvement in northeast Kansas. J Environ Manage 79, 336-347.

Okamura, K., Phillips, M.D., Tyler, D.M., Duan, H., Chou, Y.T., and Lai, E.C. (2008). The regulatory activity of microRNA* species has substantial influence

on microRNA and 3' UTR evolution. Nat Struct Mol Biol 15, 354-363.

Ori, N., Cohen, A.R., Etzioni, A., Brand, A., Yanai, O., Shleizer, S., Menda, N., Amsellem, Z., Efroni, I., Pekker, I., Alvarez, J.P., Blum, E., Zamir, D., and Eshed, Y. (2007). Regulation of LANCEOLATE by miR319 is required for compound-leaf development in tomato. Nat Genet 39, 787-791.

Park, M.Y., Wu, G., Gonzalez-Sulser, A., Vaucheret, H., and Poethig, R.S. (2005). Nuclear processing and export of microRNAs in Arabidopsis. Proc Natl Acad Sci U S A 102, 3691-3696.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G., Hellsten, U., Mitros, T., Poliakov, A., Schutz, J., Spannagl, M., Tang, H., Wang, X., Wicker, T., Bharti, A.K., Chapman, J., Feltus, F.A., Gowik, U., Grigoriev, I.V., Lyons, E., Maher, C.A., Martis, M., Narechania, A., Otillar, R.P., Penning, B.W., Salamov, A.A., Wang, Y., Zhang, L., Carpita, N.C., Freeling, M., Gringle, A.R., Hash, C.T., Keller, B., Klein, P., Kresovich, S., McCann, M.C., Ming, R., Peterson, D.G., Mehboob-ur-Rahman, Ware, D., Westhoff, P., Mayer, K.F.X., Messing, J., and Rokhsar, D.S. (2009). The Sorghum bicolor genome and the diversification of grasses. Nature. 457: 551-555

Peragine, A., Yoshikawa, M., Wu, G., Albrecht, H.L., and Poethig, R.S. (2004). SGS3 and SGS2/SDE1/RDR6 are required for juvenile development and the production of trans-acting siRNAs in Arabidopsis. Genes Dev 18, 2368-2379.

Pilcher, R.L., Moxon, S., Pakseresht, N., Moulton, V., Manning, K., Seymour, G., and Dalmay, T. (2007). Identification of novel small RNAs in tomato (Solanum lycopersicum). Planta 226, 709-717.

Rajagopalan, R., Vaucheret, H., Trejo, J., and Bartel, D.P. (2006). A diverse and evolutionarily fluid set of microRNAs in Arabidopsis thaliana. Genes Dev 20, 3407-3425.

Rajewsky, N. (2006). microRNA target predictions in animals. Nat Genet 38 Suppl, S8-13.

Reddy, A.M., Zheng, Y., Jagadeeswaran, G., Macmil, S.L., Graham, W.B., Roe, B.A., Desilva, U., Zhang, W., and Sunkar, R. (2009). Cloning, characterization and expression analysis of porcine microRNAs. BMC Genomics. 10:65

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. (2003). Asymmetry in the assembly of the RNAi enzyme complex. Cell. 115(2): 199-208

Shukla, L.I., Chinnusamy, V., and Sunkar, R. (2008). The role of microRNAs and other endogenous small RNAs in plant stress responses. Biochim Biophys Acta 1779, 743-748.

Subramanian, S., Fu, Y., Sunkar, R., Barbazuk, W.B., Zhu, J.K., and Yu, O. (2008). Novel and nodulation-regulated microRNAs in soybean roots. BMC Genomics 9, 160.

Sunkar, R., and Zhu, J.K. (2004). Novel and stress-regulated microRNAs and other small RNAs from Arabidopsis. Plant Cell 16, 2001-2019.

Sunkar, R., and Jagadeeswaran, G. (2008). In silico identification of conserved microRNAs in large number of diverse plant species. BMC Plant Biol 8, 37.

Sunkar, R., Girke, T., and Zhu, J.K. (2005a). Identification and characterization of

endogenous small interfering RNAs from rice. Nucleic Acids Res 33, 4443-4454.

Sunkar, R., Kapoor, A., and Zhu, J.K. (2006). Posttranscriptional induction of two Cu/Zn superoxide dismutase genes in Arabidopsis is mediated by downregulation of miR398 and important for oxidative stress tolerance. Plant Cell 18, 2051-2065.

Sunkar, R., Girke, T., Jain, P.K., and Zhu, J.K. (2005b). Cloning and characterization of microRNAs from rice. Plant Cell 17, 1397-1411.

Sunkar, R., Zhou, X., Zheng, Y., Zhang, W., and Zhu, J.K. (2008). Identification of novel and candidate miRNAs in rice by high throughput sequencing. BMC Plant Biol 8, 25.

Szarzynska, B., Sobkowiak, L., Pant, B.D., Balazadeh, S., Scheible, W.R., Mueller-Roeber, B., Jarmolowski, A., and Szweykowska-Kulinska, Z. (2009). Gene structures and processing of Arabidopsis thaliana HYL1-dependent pri-miRNAs. Nucleic Acids Res 37, 3083-3093.

Talmor-Neiman, M., Stav, R., Frank, W., Voss, B., and Arazi, T. (2006a). Novel micro-RNAs and intermediates of micro-RNA biogenesis from moss. Plant J 47, 25-37.

Talmor-Neiman, M., Stav, R., Klipcan, L., Buxdorf, K., Baulcombe, D.C., and Arazi, T. (2006b). Identification of trans-acting siRNAs in moss and an RNA-dependent RNA polymerase required for their biogenesis. Plant J 48, 511-521.

Thatcher, E.J., Paydar, I., Anderson, K.K., and Patton, J.G. (2008). Regulation of zebrafish fin regeneration by microRNAs. Proc Natl Acad Sci U S A 105, 18384-18389.

Tran, R.K., Zilberman, D., de Bustos, C., Ditt, R.F., Henikoff, J.G., Lindroth, A.M., Delrow, J., Boyle, T., Kwong, S., Bryson, T.D., Jacobsen, S.E., and Henikoff, S. (2005). Chromatin and siRNA pathways cooperate to maintain DNA methylation of small transposable elements in Arabidopsis. Genome Biol 6, R90.

Vaucheret, H. (2005). MicroRNA-dependent trans-acting siRNA production. Sci STKE 2005, pe43.

Vazquez, F., Gasciolli, V., Crete, P., and Vaucheret, H. (2004a). The nuclear dsRNA binding protein HYL1 is required for microRNA accumulation and plant development, but not posttranscriptional transgene silencing. Curr Biol 14, 346-351.

Vazquez, F., Vaucheret, H., Rajagopalan, R., Lepers, C., Gasciolli, V., Mallory, A.C., Hilbert, J.L., Bartel, D.P., and Crete, P. (2004b). Endogenous trans-acting siRNAs regulate the accumulation of Arabidopsis mRNAs. Mol Cell 16, 69-79.

Williams, L., Carles, C.C., Osmont, K.S., and Fletcher, J.C. (2005). A database analysis method identifies an endogenous trans-acting short-interfering RNA that targets the Arabidopsis ARF2, ARF3, and ARF4 genes. Proc Natl Acad Sci U S A 102, 9703-9708.

Willmann, M.R., and Poethig, R.S. (2007). Conservation and evolution of miRNA regulatory programs in plant development. Curr Opin Plant Biol 10, 503-511.

Wu, G., **and Poethig, R.S.** (2006). Temporal regulation of shoot development in Arabidopsis thaliana by miR156 and its target SPL3. Development. 133(18): 3539-3547

Xie, Z., Allen, E., Fahlgren, N., Calamar, A., Givan, S.A., and Carrington, J.C.

(2005). Expression of Arabidopsis MIRNA genes. Plant Physiol 138, 2145-2154.

Yang, Z., Ebright, Y.W., Yu, B., and Chen, X. (2006). HEN1 recognizes 21-24 nt small RNA duplexes and deposits a methyl group onto the 2' OH of the 3' terminal nucleotide. Nucleic Acids Res 34, 667-675.

Yu, B., **Bi, L., Zheng, B., Ji, L., Chevalier, D., Agarwal, M., Ramachandran, V., Li, W., Lagrange, T., Walker, JC., and Chen, X.** (2008). The FHA domain protien DAWDLE in *Arabidopsis* and SNIP1 in humans act in small RNA biogenesis. Proc Natl Acad Sci USA. 105, 10073-10078.

Yu, B., **and Wang, H.** (2010). Translational inhibition by microRNA in plants. Prog mol subcell biol 50, 41-57

Zhang, B., Pan, X., Cannon, C.H., Cobb, G.P., and Anderson, T.A. (2006). Conservation and divergence of plant microRNA genes. Plant J 46, 243-259.

Zhang, W., Luo, Y., Gong, X., Zeng, W., and Li, S. (2009). Computational identification of 48 potato microRNAs and their targets. Comput Biol Chem 33, 84-93.

Zhao, T., Li, G., Mi, S., Li, S., Hannon, G.J., Wang, X.J., and Qi, Y. (2007). A complex system of small RNAs in the unicellular green alga Chlamydomonas reinhardtii. Genes Dev 21, 1190-1203.

Zhong, R., **and Ye, Z.H.** (2007). Regulation of HD-ZIP III Genes by MicroRNA 165. Plant Signal Behav 2, 351-53

Zhou, G.K., Kubo, M., Zhong, R., Demura, T., and Ye, Z.H. (2007). Overexpression of miR165 affects apical meristem formation, organ polarity establishment and vascular development in Arabidopsis. Plant Cell Physiol 48, 391-404

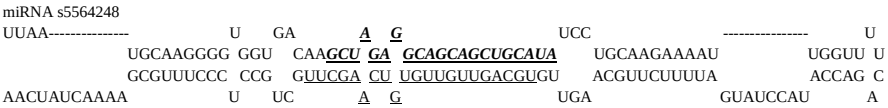
Zhu, Q.H., Spriggs, A., Matthew, L., Fan, L., Kennedy, G., Gubler, F., and Helliwell, C. (2008). A diverse set of microRNAs and microRNA-like small RNAs in developing rice grains. Genome Res 18, 1456-1465.

Zilberman, D., Cao, X., Johansen, L.K., Xie, Z., Carrington, J.C., and Jacobsen, S.E. (2004). Role of Arabidopsis ARGONAUTE4 in RNA-directed DNA methylation triggered by inverted repeats. Curr Biol 14, 1214-1220..

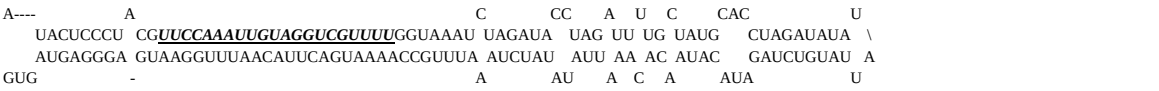
APPPENDICES

Appendix 1: Predicted fold-back structures for the novel miRNA precursors

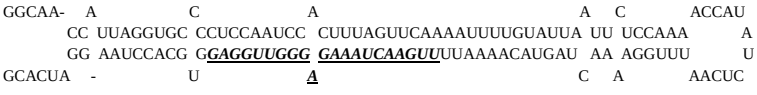
miRNA s5564248



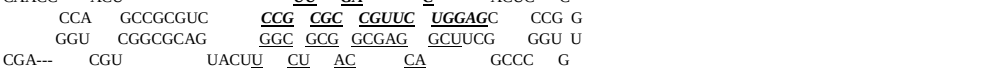
miRNA s8018588



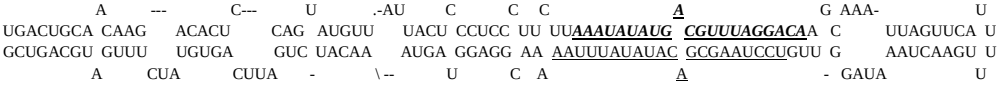
miRNA s8096486



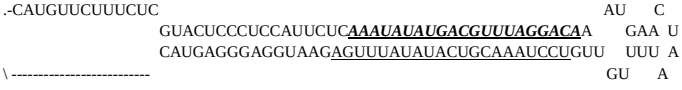
miRNA s4001019



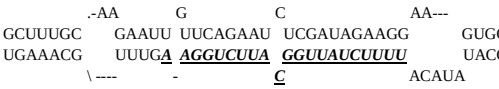
miRNA s185087_h0



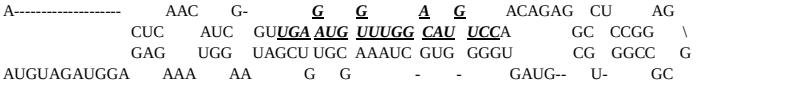
miRNA s185087_h4



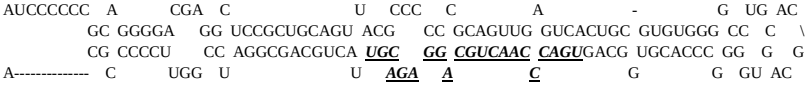
miRNA s8382548



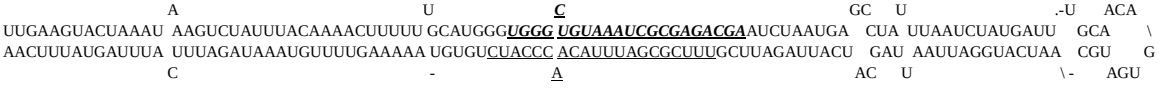
miRNA s7470078



miRNA s7437949



miRNA s7724135



miRNA s850747
GUCAUC-- UUUAU - UCAU CCA U AUU G UC U
UGCGUG CAUG UAGGG UUUGAU UGCCACUACAAUUUC UGAAGUUGG UCAUGUU AAA CAUGUCAU \
ACGCAU GUAC AUCCC AAGCUA ***ACGGUGGUGUUGAGG ACUUUAA***CC AGUGUAA UUU GUACGGUG G
CACUAGCU ---------- U UUUU AGC ***C*** CAC G UA A

miRNA s6651927
.-GCCUCC .-UCGA G ***A*** ***GG*** ***C*** -- GGAUG A U -| UCA
UCAGGAUGG CUG ***GUUGU GAGAU*** ***CUG UU GAACA*** CUC ACAU GA AUGUG \
GGUCUUACC GAC CAGCG CUUUG GAC AA CUUGU GAG UGUA CU UACAU A
\ ---------- \ ------- G A AA U U AACAA G C G^ GCG

miRNA s1509072
UU .-AUUAG -------- -- -- ***C*** ***GA*** ***CA*** ***UU*** A ----------- G
AGCUAGG UUCAAUCU AGA UGCA***AG AA ACGAU ACACAG GG*** UAG GU GGUUCGA C
UCGAUUC GAGUUGGA UCU GCGUUC UU UGUUG UGUGUC CU AUC CA CCAGGCU C
---- \ ---------- GAAU C G A --- AC GC C UAAUG G

miRNA s6650602
---------- A C C-- G A --- CU -- CU A CAUA
GCGGAC GU CGCC CAGGGGC CGGAC GUCCGCC CGGGGCU AACGGU AGA CUG CA \
CGCCUG CA GCGG GUCUCUG GC***CUG CAGGCGG GUUUUGG UUG***CCG UCU GAC GU A
AAAAG A A AGU -- ***G UAA*** ---- A U-- -- UACU

miRNA s2472577
GUAGAUCA| UCC CC UUU G UGG AU AG CUCCU U
UG GCAC AAGCU C AG AGCCGUG GGAUGA AGCUA GCAUG C
AC UGUG UUCGA ***G UC UUGGUAU CCUAUU UUGA***U CGUAU G
UC----------^ UGU U- UGU -- ***UUG*** --- ***G***- -------- U

miRNA s3495498
AA .-AACUAA AAAAUUC U CA CCG C - CU
UCUU UAGGG GAU CAUGCCAC CAACU UGAAAUUGGGUUUCAUGUU AAAAUC AUGCCA \
AGAG AUCCC CUA ***GUACGGUG GUUGA ACUUUA***ACUUAAAGUACAA UUUUAG UACGGU C
--- \ ---------- AGUCAAC U ***AG AGA*** - G AA

miRNA s3377122
ACU -- CC---- U GCCCAA -- C C-- CU CUACCGACCCA A AA
GAA ACG CUGUGGAC GUCCGC GUGGC GGAC GUCCGC GUA AUU CCA AGAC C
CUU UGC GACGCCUG CAGGCG CA***CCG CCUG CAGGCG CAU UA***A GGU UCUG A
AU- A ACUU C GUUCC- ***G U UC CU*** UCUUUAACAA- C CA

miRNA s343898
GA ***AAC***- ***U A AGG*** ----| U
GG ***GGGCG CG CGGCGUCU G***CCUC GC G
CC UCUGC GC GCUGCGGA UGGAG CG C
UG GGGA -- G CAG AG^ G

miRNA s8061400
CUGA---------- UUUUU C---- -- U A A-- U U CGCUC
AGGC ACGA GUC CAGGUUCG CGAACCUG GGC CAU AAU GC \
UCCG ***UGCU CGG GUCCAAGU GCUU***GGAC CUG GUA UUA CG A
CGACCAAAUAUA ---------- ***AUA A U*** - CA C C AAACU

miRNA s3019914
AUA -- -- UG -- UG ------ A G A- -- ACU UG UGG
UUCGUUUGGCU GU GGC AU GGC GUG UUG UUU UUAUG GAA CAGU GCUGAC CC U
AAGCAAACCGA CA UCG UG ***UCG CGU GAC AGA AGU***AU UUU GUCG UGAUUG GG A
AC-- A G GU ***G GU CAU A G*** GA A ------ GU UCG

miRNA s6532536
CG UU UCUCUC ***U A CU*** -- A ---- CAC
AU CCCCUGCG ***GUGUGC GAG CCUG GCG GGCU*** CGC CGGG U
UA GGGGAUGC CGCACG CUC GGGC CGU UCGG GCG GUCC A
G- GG ---------- -- -- ---- A -- UU AAG

miRNA s1781441
CG U CU- AUGG A UCAC UU AAC - CUACAGUAA U
GUACUAAA AAGUU AUUUGCAAAAC CUUCACGG GUGU AUUUU GACGAAUCUAAUG GGU UAAUUG AUGAUUGG UGC A
CAUGAUUU UUCAG UAAAUGUUUUG GAGGUGUU ***CGCG UGAAG CUGCUUAGA***UUAC CCG AUUGGC UACUAACC AUG C
AU U AUU GGG***A A CGCU*** U- GAA G UCCAACCA- A

miRNA s5714794
CGG***GG G GU A A*** -- GC UG G CUAC C U A
AC UG CUGUAG GU GGUG UGUC UC GA GCC GU UCUACC GG C
UG AU GGCGUC CA CUAU AUAG AG CU UGG CA GGGUGG CC G
AGAUG G UG -- G G A- GU G A----- C -- U

miRNA s1951091

```
    CU    U       --        --    U-   CU        U-   GAGA      A    .-UGUA    C--|    AG   GAG
GGG   GGC  GGAGC  UGAAGC  UGG   GG   GGAGUGG   GU     GAAAA  UAC      GCG   UGG   GG    \
UCC   UCG  CCUUG  ACUUCG  GCC   UC   UCUUGCC   CA     UUUUU  AUG      CGC   GCC   CC    A
   U-    U       U        U   UC   UU         UC   AAG-       G   \--------    AA^   GA   AGU
```

miRNA s1901506

```
G-------------------------------------------------                U     --       CA--          GU       C
                                                    GGCUCCGGCUCU  CCA  AAAA     GCUCCG   UCUGG  U
                                                    CCGAGGUUGAGG  GGU  UUUU     CGAGGU   AGACU  U
ACAAAACGGUUUAUAAAAGAUUUUG                           U     C      CGG          ----       U
```

miRNA s8008250

```
-                                                                                                                  GG          ACA
 CAUGGCUUCACCUUGCCUGUGUAAGUCGCUGAUGGCUUCAGGACCGGCUUCACACGUGAAGCC   UCCCAAA   \
 GUACCGAAGUGGAACGGACACAUUCAGCGACUACCGAAGUCCUGGCCGAAGUGUGCACUUCGG   AGGGUUU    A
A                                                                                                                  GG          GCA
```

miRNA: s6815382

```
U-        C       U                          C     AA    A       A-        UGCUAC|  U       CUAC
   GUAA   UUG  UGAGAUGAAUCUAAUGA  GGU   UU  AUUG   UGAUU           AG GAUG        A
   CAUU   AGC  GCUCUACUUAGAUUACU  CCG   AA  UGGC   ACUAA           UC CUAC        A
CA       U      --                    --     GA   C       GC          UU--------^   --       CAAU
```

miRNA s6703270

```
                    CA       G                                          UG     A           -   A     A     G         --      U
AUAAAACUUUUUGUA  GAUGG  UUGUAAAUCGCGAGACGAAUCUAAUGA   CUA  UUAAUCC  AU  AUU  GCA AUGGU    UAC G
UGUUUUGAAAAACGU   CUACC  GACAUUUAGCGCUCUGCUUAGAUUACU   GAU  AAUUAGG UA  UAA  CGU UGUCA    AUG U
                        AA        A                                          UG     G           G   C     A     -         AC      A
```

miRNA s3496977

```
A----          UC       G     C                        AC             A                  UC     A      U
     GAAGGG   AGUU  GAU  CAUGCCACUACAAUUUC   GAAGUUGG  UUUCAUGUUGAAA   CAU CCAU  \
     CUUCCC   UUAA  CUA  GUACGGUGGUGUUGAGG   CUUUAACC  AAAGUGCAACUUU   GUA GGUG  G
AAC          UU       G     A                        CA             C                  UA     C      A
```

miRNA s3496977

```
U---                  UC       G     C            UC           U              AUC                              U
    UUUAUAAGGGG   AGUU  GAU  CAUGCCAC   CAAUUUC  UGAAAUUGG   UUAUGUUGAAAAUCAUGCCAU  \
    GAAUAUUUCCC   UUAA  CUA  GUACGGUG   GUUGAGG  ACUUUAACC   AGUACAACUUUUAGUACGGUG  G
ACC                   UU       G     A            GU           C              CAC                              A
```

Appendix 2: Alignments between conserved miRNAs and their predicted targets

```
EXTA9310.b1 (SPL8)          606  ATGCTCTCTCTCTTCTGTCA  625
                                  ||||||||||||||||||||
     MIR156                 21   TACGAGAGAGAGAAGACAGT  1

CBYY6990.b1 (SPL like)      536   GTGCTCTCTCTCTTCTGTCA  555
                                  :||||||||||||||||||||
     MIR156                 21   TACGAGAGAGAGAAGACAGT  1

FL921090 (AT-GTL1)          439  GGCAGCTCCGGGGGCATGCAA   458
                                  0||||||||0||||||||||
MiR159                      21   ACGTCGAGGACCCCGTACGTT    1

CBYZ12350.g1(MYB31)         371   AGCAGCTCCTTTCAATCCAAA  391
                                  |:0||||||||||||||||||
MiR159e                      21  TTCTCGAGGAAAGTTAGGTTT  1

GD051711 (Hypothetical protein):
                        224 CAGAGCTCCCTTCAATCCAAA 244
                            ||||||||||||||||||||
MIR159:                  21 GTCTCGAGGGAAGTTAGGTTT 1

GD037618 (Unknown protein):
                           137 CAGAGCTCCCTTCAATCCAAA 157
                               ||||||||||||||||||||
MIR159:                     21 GTCTCGAGGGAAGTTAGGTTT  1

FL997283 (Expressed protein):
                        190 CAGAGCTCCCTTCAATCCAAA 210
                            ||||||||||||||||||||
MIR159:                  21 GTCTCGAGGGAAGTTAGGTTT 1

FL913173 (ARF18)             510 AGGCATACAGGGAGCCAGGCA 530
                               0||||||||||||||||||||
MIR160                        21  ACCGTATGTCCCTCGGTCCGT  1

FL738979 (hypothetical protein)
                              348 TGGCATACAGGGAGCCAGGCA 328
                                  ||||||||||||||||||||
MIR160                         21  ACCGTATGTCCCTCGGTCCGT 1
```

FE606478 (START domain containing protein)

```
                          606 CTGGGATGAAGCCTGGTCCGG 626
                              ||||||||||||||||||||
MIR160c:                   21 GACCCTACTTCGGACCAGGCC  1
```

CCGG13080.b1 (DCL1)

```
                         249  CTGGATGCAGAGGTTTTATCG 269
                              ||||||||||||||||00:0:
MIR162                    21  GACCTACGTCTCCAAATAGCT 1
```

4855_1_CBYX_CBYY_CBYZ_EXTA (NAC TRANSCRIPTION FACTOR 9)

```
                         754  AGCAAGTGCCCTGCTTCTCCA 774
                              0|||0|||||||||||||||
MIR164                    21  AGCTGCACGGGACGAAGAGGT 1
```

FL8462281 (NAM)

```
                    59    AGCTCGTGCCCTGCTTCTCCA 79
                          0||0||||||||||||||||
MiR164              21    ACGTGCACGGGACGAAGAGGT 1
```

9905_0_CBYX6738.b1_CBYX_CBYY_CBYZ_EXTA (REV)

```
                         606  CTGGGATGAAGCCTGGTCCGG 626
                              ||||||||||||||||||||
MiR166c                   21  GACCCTACTTCGGACCAGGCC 1
```

GD002178 (HEAT repeat family protein):

```
                       567 AGGGGAATGAAGCCTGGTCCGA 588
                           ||||||||||||||||||||||
MIR166:                 22 TCCCCTTACTTCGGACCAGGCT 1
```

FL954559 (Unknown protein):

```
                         694 AGGGGAATGAAGCCTGGTCCGA 715
                             ||||||||||||||||||||||
MIR166:                   22 TCCCCTTACTTCGGACCAGGCT 1
```

GD032712 (Unknown protein):

```
                          404 CCAGATCATGCTGGCAGCTTCA 425
                              ||||||||||||||||||||||
MIR167:                    22 GGTCTAGTACGACCGTCGAAGT 1
```

GD007307 (Hypothetical protein):

```
                       739 CCAGATCATGCTGGCAGCTTCA 759
                           ||||||||||||||||||||||
MIR167:                 22 GGTCTAGTACGACCGTCGAAGT 1
```

FL904157(Argo1): 224 TTCCCGAGCTGCACCAAGCCC 244
|||||0||||||||||||0
MIR168 21 AAGGGCACGACGTGGTTCGGT 1

FL818096 (PINHEAD) 164 TTCCCGAGCTGCACCAAGCCC 184
|||||0||||||||||||0
MIR168 21 AAGGGCACGACGTGGTTCGGT 1

FL965734 (HAP2C) 449 GCGGCAATTCATCCTTGGCTT 469
0||||||0||||||||||||0
MiR169 21 AGCCGTTCAGTAGGAACCGAC 1

FL910918 (putative scarecrow transcription factor):
547 AGATATTGGCGCGGCTCAATCA 567
||||||||||||||||||||||
MIR171: 22 TCTATAACCGCGCCGAGTTAGT 1

6988_1_CBYX_CBYY_CBYZ_EXTA (Scarecrow like transcription factor 9)
66 AGATATTGGCGCGGCTCAATTA 87
||||||||||||||||||||:|
MiR171g 22 TCTATAACCGCGCCGAGTTAGT 1

CBYZ4142.b1 (TOE2)
380 CTGCAGCATCATCAGGATTCT 400
|||||||||||||:||||||
MiR172 21 GACGTCGTACTAGTTCTAAGA 1

FL945982 (AP2 domain containing protein):
503 CTGCAGCATGATGAGGATTCT 523
|||||||||||||:||||||
MIR172: 21 GACGTCGTACTAGTTCTAAGA 1

FL945982 (AP2 domain containing protein):
89 CTGCAGCATGATGAGGATTCT 109
|||||||||||||:||||||
MIR172: 21 GACGTCGTACTAGTTCTAAGA 1

1618_1_CBYX_CBYY_CBYZ_EXTA (TCP24)
246 AGGGGGACCCTTCAGTCCAA 266
0||:|0||||||||||||||
MiR319 20 CCCTCGTGGGAAGTCAGGTT 1

FL985594 (Expressed protein):

```
                          536 GGGAGCACCCTTCAGTCCAA 555
                              ||||||||||||||||||||
MIR319:                    20 CCCTCGTGGGAAGTCAGGTT 1
```

FL692881 (Leucine rich repeat protein):

```
                          298 GGCGCTATGCCTCCTGAGCTT 318
                              ||||||||0|||||||||||
MIR390                     21 CCGCGATAAGGAGGACTCGAA 1
```

0_0_EXTA9190.b1_CBYX_CBYY_CBYZ_EXTA (AFB2)

```
                          231   AGACAATGCGATCCCTTTGGA 251
                                0:0||||||||||||||||||
MiR393                    21    CTAGTTACGCTAGGGAAACCT 1
```

FL978450 (F-BOX containing protein):

```
                          272 GGAGGTGGACAGAATGCCAA 291
                              ||||||||||||||||||00
MIR394:                    20 CCTCCACCTGTCTTACGGAA 1
```

FL910325 (bifunctional 3’-phosphoadenosine 5-phosphosulfate synthase):

```
                          421 GAGTTCCTCCAAGCACTTCAT 441
                              ||||||||||||:||||||||
MIR395:                    21 CTCAAGGAGGTTTGTGAAGTA 1
```

FL710917 (Sulfate transporter)

```
                          39  GAGTTCCCCCAAACACTTCAC  59
                              ||||||||||||0||||||||
MIR395                    21  CTCAAGGGGGTTCGTGAAGTG   1
```

FL753322 (Laccase precursor protein):

```
                          436 CATCAACGCTGCACTCAATGA 456
                              ||||||||||||||||||||
MIR397:                    21 GTAGTTGCGACGTGAGTTACT 1
```

FL997840 (Transposon protein):

```
                          595 CAGGGCAATTCTCCTTTGGCA 615
                              ||||||||||||||||||||
MIR399:                    21 GTCCCGTTAAGAGGAAACCGT 1
```

FL811879 (Ubiquitin conjugating enzyme protein)

```
                          518 CAGGGCAAATCTCCTTTGGCT 538
                              |0||||||||||||||||0
MIR399:                    21 GACCCGTTTAGAGGAAACCGT 1
```

FL942386 (Plastocyanin like domain containing protein):

```
                      388 GCCAGGGTAGAGGCAGTGCAG 408
                          |||||||0|||||||||||||
MIR408:                 21 CGGTCCCTTCTCCGTCACGTC 1
```

FL897156 (BBTI12)

```
                        3     AAGTCAAACTTCTCTAAGTCC 23
                              ||||||||||||||||||0|00
MiR437                  21    TTCAGTTTGAAGAGATTGAAA 1
```

FL979804 (MADS box family protein):

```
                       466 AAGCTTGAGGCAGCAACTGCA 486
                           |||||||||||||||||||||
MIR444:                 21 TTCGAACTCCGTCGTTGACGT 1
```

GD052089 (Expressed protein):

```
                        482 CTCCTCTGCATGCCCCTTCCA 512
                            |||||||||||||||||||||
MIR528:                  21 GAGGAGACGTACGGGGAAGGT 1
```

Appendix 3: Alignments between novel microRNAs and their predicted targets

FL788822 (Hypothetical protein)

```
110   UUCACGUGUGAAGCUGGUCCUGAA  133    FL788822
      |||||||||||||:|||||||||
3'    AAGUGCACACUUCGGCCAGGACUU   5'    s8008250
```

FL695873 (UDP-glycoronsyl/UDP-glycosyl transferase protein)

```
135   UUCACGUGUGAAGCCGGUUCUGAA   158   FL695873
      ||||||||||||||||||:|||||
3'    AAGUGCACACUUCGGCCAGGACUU   5'    s8008250
```

GD043417 (MTA/SAH nucleosidase)

```
362   UUCAUGUGUGAAGCCGGUUCUGAA   285   GD043417
      ||||:|||||||||||||:|||||
3'    AAGUGCACACUUCGGCCAGGACUU   5'    s8008250
```

GD016940 (Sulfotransferase domain containing protein)

```
402   AUCAUCAAUUAAUUACCGUCAUUA   425   GD016940
      ||||||||||||||||||||||||
3'    UAGUAGUUAAUUAAUGGCAGUAAU   5'    s6815382
```

FL961655 (Expressed protein)

```
43    UCGUCUCGCGAUUUACAACCCAUC   66    FL961655
      ||||||||||||||||||||||||
3'    AGCAGAGCGCUAAAUGUUGGGUAG   5'    s6703720
```

FL960500 (Nucleolar RNA binding protein Nop10p like)

```
36    UCGUCUCGCGAUUUACAACCCAUC   59    FL960500
      ||||||||||||||||||||||||
3'    AGCAGAGCGCUAAAUGUUGGGUAG   5'    s6703720
```

FL938010 (putative dehydrogenase protein)

```
390   UCGUCUCGCGAUUUACAACCCAUC   413   FL938010
      ||||||||||||||||||||||||
3'    AGCAGAGCGCUAAAUGUUGGGUAG   5'    s6703720
```

FL993113 (amino acid permease/amino acid transporter)

```
121   CAUGCCACCACAACUCCGUGAAAU   144   FL993113
      ||||||||||||||||||||||||
3'    GUACGGUGGUGUUGAGGCACUUUA   5'    s3496977
```

FL987523 (Hypothetical protein)

```
114   CAUGCCACCACAACUCCGUGAAAU   137   FL987523
      ||||||||||||||||||||||||
3'    GUACGGUGGUGUUGAGGCACUUUA   5'    s3496977
```

FL967210 (Expressed protein)

```
217   CAUGCCACCACAACUCCGUGAAAU   240   FL967210
      ||||||||||||||||||||||||
3'    GUACGGUGGUGUUGAGGCACUUUA   5'    s3496977
```

FL846993 (Hypothetical protein)

```
116   CAUGCCACCACAACUCCGUGAAAU   139   FL846993
      ||||||||||||||||||||||||
3'    GUACGGUGGUGUUGAGGCACUUUA   5'    s3496977
```

FL821059 (Protease inhibitor, seed storage, LTP protein)

```
711   CAUGCCACCACAACUCCGUGAAAU   734   FL821059
      ||||||||||||||||||||||||
3'    GUACGGUGGUGUUGAGGCACUUUA   5'    s3496977
```

FL754750 (Splicing factor U2AF protein)

```
257   CAUGCCACCACAACUCCGUGAAAU   280   FL754750
      ||||||||||||||||||||||||
3'    GUACGGUGGUGUUGAGGCACUUUA   5'    s3496977
```

FL870890 (Hypothetical protein)

```
590   CUCACACCACUCCAGCCACCAGCU   613   FL870890
      ||||||||||||||||||||||||
3'    GAGUGUGGUGAGGUCGGUGGUCGA   5'    s1951091
```

FL847596 (Hypothetical protein)

```
467   CUCACACCACUCCAGCCACCAGCU   490   FL847596
      ||||||||||||||||||||||||
3'    GAGUGUGGUGAGGUCGGUGGUCGA   5'    s1951091
```

GD025927 (Activator of 90kDa heat shock protein ATPase)

```
23    UCCAACUCCACCAGAAAAGCCGCU   46    GD025927
      ||||||||||||||||||||||||
3'    AGGUUGAGGUGGUCUUUUCGGCGA   4'    s1901506
```

GD038932 (Hypothetical protein)

```
77    CAGCCAGCAGUACUGUUCUCUCAU   100   GD038932
      ||||||||||||||||||||||||
3'    GUCGGUCGUCAUGACAAGAGAGUA   5'    s3019914
```

FL897758 (Hypothetical protein)

```
522   CAGCCAGCAGUACUGUUCUCUCAU   545   FL897758
      ||||||||||||||||||||||||
3'    GUCGGUCGUCAUGACAAGAGAGUA   5'    s3019914
```

FL774345 (Unknown protein)

```
234   CAGCCAGCAGUACUGUUCUCUCAU   257   FL774345
      ||||||||||||||||||||||||
3'    GUCGGUCGUCAUGACAAGAGAGUA   5'    s3019914
```

FL768681 (Hypothetical protein)

```
54    CAGCCAGCAGUACUGUUCUCUCAU    77    FL768681
      ||||||||||||||||||||||||
3'    GUCGGUCGUCAUGACAAGAGAGUA    5'    s3019914
```

FL892897 (Hypothetical protein)

```
587   AGCCCGCAGCAGGUCUCAGCACAC    610   FL892897
      ||||||||||||||||||||||||
3'    UCGGGCGUCGUCCAGAGUCGUGUG    5'    s6532536
```

FL854510 (Hypothetical protein)

```
413   UGCGCUACUUCACGAGACGAAUCU    436   FL854510
      ||||||||||||0||||||||||||
3'    ACGCGAUGAAGCGCUCUGCUUAGA    5'    s1781441
```

GD036874 (Pantothenate kinase)

```
70    CAGAACAACCAUAGGAUAACAACU    93    GD036874
      ||||||||||||||||||||||||
3'    GUCUUGUUGGUAUCCUAUUGUUGA    5'    s2472577
```

FL787755 (Expressed protein)

```
566   CAGAACAACCAUAGGAUAACAACU    589   FL787755
      ||||||||||||||||||||||||
3'    GUCUUGUUGGUAUCCUAUUGUUGA    5'    s2472577
```

FL891811 (Hypothetical protein)

```
23    CAUGCCACUCCAACUUCUUGAAAU    46    FL891811
      ||||||||||||||||||||||||
3'    GUACGGUGAGGUUGAAGAACUUUA    5'    s3495498
```

FL821060 (Hypothetical protein)

```
82    CAUGCCACUCCAACUUCUUGAAAU    104   FL821060
      ||||||||||||||||||||||||
3'    GUACGGUGAGGUUGAAGAACUUUA    5'    s3495498
```

FL952548 (Hypothetical protein)

```
279   CAUGCCACUCCAAAUUCUUGAAAU    302   FL952548
      ||||||||||||0||||||||||||
3'    GUACGGUGAGGUUGAAGAACUUUA    5'    s3495498
```

FL892558 (Hypothetical protein)

```
257   CAUGCCACUCCAAAUUCUUGAAAU    280   FL892558
      ||||||||||||0||||||||||||
3'    GUACGGUGAGGUUGAAGAACUUUA    5'    s3495498
```

FL858647 (Hypothetical protein)

```
337   CCCUAGACGCCGUCGACGCCCGUU    360   FL858647
      ||||||||||||||||||||||||
3'    GGGAUCUGCGGCAGCUGCGGGCAA    5'    s343898
```

FL856169 (Hypothetical protein)

```
247   CCCUAGACGCCGUCGACGCCCGUU   270   FL856169
      ||||||||||||||||||||||||
3'    GGGAUCUGCGGCAGCUGCGGGCAA   5'    s343898
```

FL804352 (Hypothetical protein)

```
349   CCCUAGACGCCGUCGACGCCCGUU   372   FL804352
      ||||||||||||||||||||||||
3'    GGGAUCUGCGGCAGCUGCGGGCAA   5'    s343898
```

FL962901 (Aldehyde oxidase)

```
494   GGACAUGUCCAAACCAUCUCA 514   FL962901
      |||||||||||||||||||||
3'    CCUGUACAGGUUUGGUAGAGU 5'    S7470078
```

FL868527 (Hypothetical protein)

```
583   ACGUCUCCUGCAGUUGGGUCA 603   FL868527
      |||||||||||||||||||||
3'    UGCAGAGGACGUCAACCCAGU 5'    s7347949
```

FL868007 (peptide transporter PTR2)

```
9     UCGUCUCGCGAUUUACAGCCCA     30    FL868007
      ||||||||||||||||||||||
3'    AGCAGAGCGCUAAAUGUCGGGU     5'    s7724125
```

FL801297 (Leucine rich repeat protein)

```
24    UCGUCUCGCGAUUUACAGCCCA     44    FL801207
      ||||||||||||||||||||||
3'    AGCAGAGCGCUAAAUGUCGGGU     5'    s7724125
```

GD035751 (Unknown protein)

```
74    UCGUCUCGCGAUUUACAACCCA     95    GD035751
      ||||||||||||||||0||||
3'    AGCAGAGCGCUAAAUGUCGGGU     5'    s7724125
```

GD019762 (Hypothetical protein)

```
184   AACCUGCUGUGUUCAUCGUCUUCU   204   GD019762
      ||||||||||||||||||||||||
3'    UUGGACGACACAAGUAGCACAAGA   5'    s1509072
```

FL963020 (Hypothetical protein)

```
314   AACCUGCUGUGUUCAUCGUCUUCU   337   FL963020
      ||||||||||||||||||||||||
3'    UUGGACGACACAAGUAGCACAAGA   5'    s1509072
```

FE657239 (Glyscosyl transferase family 1 like protein)

```
167   UAUGCAGCUGCUGCCUCUAGC 187   FE657239
      |||||||||||||||||||||
3'    AUACGUCGACGACGGAGAUCG 5'    s5564248
```

FL711513 (WD40 repeat protein)

```
37    AAAAGCACCUACAAUUUGGAA 57    FL711513
      ||||||||||||||||||||
3'    UUUUCGUGGAUGUUAAACCUU 5'    s8018588
```

FL787859 (Exportin 1)

```
719   CUCCAACCCUCUUUAGUUCAA 739   FL787859
      ||||||||||||||||||||
3'    GAGGUUGGGAGAAAUCAAGUU 5'    s8096486
```

GD022328 (Expressed protein)

```
378   UGUCCUAAACGUCAUAUAUUU 398   GD022328
      ||||||||||||||||||||
3'    ACAGGAUUUGCAGUAUAUAAA 5'    s185087
```

FL978160 (Acyl-CoA dehydrogenase, mitochondrial precursor)

```
602   UGUCCUAAACGUCAUAUAUUU 621   FL978160
      ||||||||||||||||||||
3'    ACAGGAUUUGCAGUAUAUAAA 5'    s185087
```

FL871007 (TUDOR protein with multiple SNc domains)

```
475   UGUCCUAAACUUCAUAUAUUU 495   FL871007
      ||||||||||0||||||||||
3'    ACAGGAUUUGCAGUAUAUAAA 5'    s185087
```

VITA

Jessica Allyn-Brooker Matts

Candidate for the Degree of

Doctor of Philosophy

Thesis: IDENTIFICATION OF MICORNAS AND THEIR TARGETS IN SWITCHGRASS, A MODEL CELLULOSIC BIOFUEL PLANT SPECIES

Major Field: Biochemistry and Molecular Biology

Biographical:

Education:

Completed the requirements for the Doctor of Philosophy in Biochemistry and Molecular Biology at Oklahoma State University, Stillwater, Oklahoma in December, 2010.

Completed the requirements for the Bachelor of Science in Biochemistry and Molecular Biology at Oklahoma State University in 2005.

Completed the requirements for the Bachelor of Science in Biology at Oklahoma State University in 2005.

Experience:

Graduate Research Assistant, Department of Biochemistry and Molecular Biology. 2005-2010. Oklahoma State University
Graduate Teaching Assistant, Department of Biochemistry and Molecular Biology. 2006-2007. Oklahoma State University
Undergraduate Research Assistant, Department of Biochemistry and Molecular Biology. 1998-2005. Oklahoma State University

Professional Memberships:

2010-present: Student Member of RNA society
2010-present: Student Member of American Society for Biochemistry and molecular biology
2005-2010: Biochemistry and Molecular Biology Graduate Student Association (member, secretary, vice president)
2003-present: Phi Lambda Upsilon, Graduate Chemistry Honor Society (member)

Name: Jessica Allyn-Brooker Matts Date of Degree: December, 2010

Institution: Oklahoma State University Location: Stillwater, Oklahoma

Title of Study: IDENTIFICATION OF MICRORNAS AND THEIR TARGETS IN SWITCHGRASS, A MODEL CELLULOSIC BIOFUEL PLANT SPECIES

Pages in Study: 95 Candidate for the Degree of Doctor of Philosophy

Major Field: Biochemistry and Molecular Biology

Scope and Method of Study:

Over the past several years, several plant species such as switchgrass, *Miscanthus*, *Sorghum*, and *Brachypodium* have been recognized as potential model plant species for cellulosic bioenergy production. Of these, switchgrass has attracted more attention in the United States, because it can be grown on marginal and wastelands, and it can also tolerate drought and heat stress. Very little is known about the basic biology of the traits that control these important characteristics, including biomass accumulation in switchgrass. Recently discovered miRNAs play an important role in post-transcriptional gene regulation and this regulation is critical for normal plant growth and development, and tolerance to environmental stress conditions including nutrient deprived conditions. To gain an insight into the complex post-transcriptional regulatory network operating in this plant species, we sought to identify the miRNAs and the genes that these miRNAs are regulating in switchgrass.

Findings and Conclusions:

By deep sequencing small RNA libraries from switchgrass, 31 conserved miRNA families, one tasiRNA family, and 15 novel miRNA families have been identified. Interestingly six of the novel miRNAs appears to be conserved in related monocot species. Small RNA blot analysis indicated that some of the conserved and novel miRNAs are expressed in a tissue-specific manner, although most are ubiquitously expressed. Surprisingly, unlike in *Arabidopsis* and other plants, miR395 and miR399 expression levels were not regulated in response to sulfate or phosphate-deprived conditions. Thirty-seven genes are predicted as targets for miRNAs, and several mRNAs (*Squamosa promoter binding-like factor, apetala2 like, NAC domain containing transcription factor*, and *HD-ZIP homologs)* were validated using 5'-RACE assays. Additionally, 45 genes are predicted as targets for novel miRNAs in switchgrass. Identification of large set of miRNAs and their targets laid the foundation for functional genomic approaches in switchgrass.

ADVISER'S APPROVAL: Dr. Ramanjulu Sunkar